DOCUMENT RESUME

ED 159 662 CS 004 427

AUTHOR Shoben, Edward J.; And Others
TITLE Issues in Semantic Memory: A Response to Glass and Holyoak. Technical Report No. 101.
INSTITUTION Bolt, Beranek and Newman, Inc., Cambridge, Mass.; Illinois Univ., Urbana. Center for the Study of Reading.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.; Public Health Service (DHEW), Arlington, Va.; Spencer Foundation, Chicago, Ill.
PUB DATE Aug 78
CONTRACT 400-76-0116
GRANT MH-19705
NOTE 85p.

EDRS PRICE MF-$0.83 HC-$4.67 Plus Postage.
DESCRIPTORS *Cognitive Processes; Comparative Analysis; Comprehension; *Language Research; *Linguistic Theory; *Memory; *Models; Psychological Studies; *Semantics
IDENTIFIERS *Center for the Study of Reading (Illinois)

ABSTRACT

Glass and Holyoak (1975) have raised two issues related to the distinction between set-theoretic and network theories of semantic memory, contending that: (a) their version of a network theory, the Marker Search model, is conceptually and empirically superior to the Feature Comparison model version of a set-theoretic theory; and (b) the contrast between set-theoretic and network theories parallels distinctions in formal semantics that are concerned with analyticity and binary truth values. This paper takes issue with both of these claims, first arguing that the set-theoretic vs. network distinction is orthogonal to issues like analyticity and binary truth. The Marker Search and Feature Comparison models are then considered in detail. Objections are raised to some of the theoretical mechanisms postulated in the Marker Search model, and Glass and Holyoak's criticisms of the Feature Comparison model are discussed. Finally, new experimental results that undermine the critical empirical base for the Marker Search model are presented. (Author/TJ)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

-ED159662

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

CENTER FOR THE STUDY OF READING

Technical Report No. 101

ISSUES IN SEMANTIC MEMORY: A RESPONSE TO GLASS AND HOLYOAK

Edward J. Shoben
University of Illinois at Urbana-Champaign
Lance J. Rips
University of Chicago
Edward E. Smith
Stanford University

August 1978

University of Illinois
at Urbana-Champaign
51 Gerty Drive
Champaign, Illinois 61820

Bolt Beranek and Newman Inc.
50 Moulton Street
Cambridge, Massachusetts 02138

The research reported in this paper was supported by U.S. Public Health Service Grant MH-19705 and a grant from the Spencer Foundation, and in part by the National Institute of Education under Contract No. US-NIE-C-400-76-0116. We thank John Anderson and Lynne Reder for making their data available to us.

CS 004427

Issues in Semantic Memory: A Response to Glass and Holyoak[1]

In their paper on "Alternative Conceptions of Semantic Memory," Glass and Holyoak (1975) raise a number of important issues concerning the psychological representation of meaning.[2] Many of these issues revolve around a distinction between set-theoretic and network models (Rips, Shoben & Smith, 1973), where the former class of models treats concepts as sets of semantic elements, while the latter class represents concepts as nodes within a network of labeled relations. With regard to this distinction, the major points of Glass and Holyoak seem to be: (1) Network models may be superior to set-theoretic ones, as suggested by a comparison of a specific set-theoretic formulation, namely the Feature Comparison model (Smith, Shoben & Rips, 1974), with a specific network proposal, the Marker Search model (Glass and Holyoak); (2) This alleged superiority of network models has definite implications for a number of well-known issues in the study of formal semantics--such as whether the distinction between analytic and synthetic truths is viable--because the set-theoretic vs. network dichotomy is intimately related to these distinctions.

We wish to challenge both of these conclusions. In the next section of this paper, we will argue that the "set-net" distinction is basically orthogonal to issues in formal semantics like the distinction between analytic and synthetic truth. We will then go on to propose a different sort of taxonomy of semantic memory models. In the third section, we will examine in detail

Glass and Holyoak's contention that the Marker Search model is superior to the Feature Comparison model. We will first offer some criticisms of the general characteristics of the Marker-Search model, and then address ourselves to some of the criticisms that Glass and Holyoak have made of the Feature Comparison model. In the fourth section, we will consider the experiments of Holyoak and Glass that, on the face of it, provide critical disconfirmations of the Feature Comparison model. Here we will present some new experimental findings that seriously qualify the Holyoak and Glass results and lessen some of the major empirical problems of the Feature Comparison model. A final section provides a summary and discussion of future directions.

The Set-Net Distinction Reconsidered

What the Distinction is Not About

Representational differences. In surveying the semantic memory literature in 1973 (Rips et al.), we found that a single representational distinction seemed to capture many of the fundamental differences among contemporary models. Thus the models proposed by Schaeffer and Wallace (1970) and Meyer (1970) had a set-theoretic structure, while the theories of Collins and Quillian (1969) and Rumelhart, Lindsay, and Norman (1972) used a network of labeled relations to represent meaning. Updating this list, one would add the Feature Comparison model as another example of a set model, and HAM (Anderson & Bower, 1973) and the Marker Search model as new instances of network models. But while this distinction served an organizational purpose, it soon became clear that the contrast between sets and nets might be a

relatively superficial indicator of more important underlying differences. This point was demonstrated by Hollan (1975) who simply noted that set models can be recast as networks by connecting each element of the set to a common node standing for the set itself.

What then can be said about our original partition of models? As we have argued elsewhere (Rips, Smith & Shoben, 1975), we still believe this partition is useful since the set-net distinction correlated with some important substantive differences among models. The task now becomes one of specifying these differences. Glass and Holyoak, who accept our distinction, have proposed two possibilities. One is that set models have considered rather simple representations and have not specified any relations among the meaning components within a concept; in contrast, network theories are capable of positing representations that stipulate entailment relations among a concept's components. We do not wish to deal at length with this proposal, but two points merit comment. Set models do not necessarily have to assume simple semantic representations, and indeed we have introduced some additional structure into set-theoretic representations (Smith, Rips & Shoben, 1974). Similarly, while network models are capable of stipulating entailment relations among meaning components, not all network models inevitably do so, as witnessed by aspects of Anderson and Bower's HAM model. Thus we think this distinction is of limited value in capturing the substantive differences between set and network theories.

Analyticity and formal vs. psychological semantics. Of greater concern to the present paper is the second distinction proposed by Glass and Holyoak.

This is their claim that set models are consistent with the view that category membership and sentence truth value are continuous or graded, while network models view category membership and truth as dichotomous. Given this assumption, Glass and Holyoak proceed to align set models with both Lakoff's (1972) advocacy of fuzzy semantics and Quine's (1953) skepticism regarding the distinction between analytic and synthetic truth; net models, in contrast, are seen as consistent with Katz's (1972) defense of analyticity and of two-valued truth. We disagree. As we see it, these important distinctions from formal semantics--analytic vs. synthetic truth and binary vs. graded truth--may be orthogonal to those substantive psychological differences that exist between the theories we have classified as set and network models.

The distinction between analytic and synthetic statements comes from philosophical semantics, and it is based on the relations among meaning entities. A statement may be classified as analytic if the meaning associated with the predicate is contained in that of the sentence subject, as in A bachelor is unmarried. Otherwise, the statement must be classified as synthetic. The analytic/synthetic distinction, then, rests on the nature of meanings and their interrelations, and not in any direct way on psychological representations. To make this point clearer, consider Frege's (1892) distinctions among the sense, reference, and idea of a word. While the sense of a word is some abstract meaning entity, its reference is the set of real-world entities denoted by the word, and its idea is roughly the psychological representation of the word. Clearly the referent and psychological representation of a word are distinct, since psychological representations are by

nature internal. Similarly, psychological representations cannot be equated with senses either, at least on Frege's account. To use Frege's own analogy, the sense of, say, moon is independent of anyone's representation of the moon in the same way that the optical image of the moon in a telescope is independent of observers' retinal images. By nature, then, theories of psychological semantics must deal primarily with individuals' representations of meanings, and not with the referents or senses themselves (Smith, Rips & Shoben, 1974).

This triptych of reference, sense, and representation has implications for a number of the arguments made by Glass and Holyoak. First, we can reject their claim that our Feature Comparison model, as presently stated, is concerned with referential meaning. This point has no force at all since our model is clearly about representations, not referents. (Indeed it is difficult to imagine how any psychological model could be solely concerned with reference.) Second, we can question their assertion that the Marker Search model, unlike the Feature Comparison model, "...is directly concerned only with sense relations" (p. 335). While psychologists may try to construct representations that capture only sense relations, current semantic-memory models, including the Marker Search model, have not done this. For example, Glass and Holyoak have used their model to explain the confirmation of sentences like Some women are writers, and such sentences clearly cannot be verified by a consideration only of sense relations, on anyone's account of sense. That is, the truth of our sample sentence is surely an empirical matter, for there is nothing about the abstract meanings of women and writers that prohibits the sentence from being false, and it is easy to imagine a set of circumstances that would make this very sentence a false one. Third,

our distinction among reference, sense, and representation allows us to make some general points about analyticity and dichotomous truth in formal vs. psychological semantics. For example, work in formal semantics indicates that the truth value of a sentence can be determined by means of relations between expressions of a language and their referents without mention of psychological representations (Tarski, 1956); it is therefore possible to adopt a binary truth-value system without implying that the psychological representations of these truth-values are also necessarily binary. In principle, then, one can endorse binary truth values in formal semantics, and continuous truth in psychological semantics. In a similar way, since the notion of analyticity can be defined in terms of the relations between the senses of expressions in a sentence, without mention of psychological factors, one can accept the analytic/synthetic distinction without implying that such a distinction need be psychologically represented. In short, the questions of whether truth values are binary and whether the analytic/synthetic distinction is tenable may be ontological questions, not psychological ones.

A more psychological approach to analyticity. While the tenability of the analytic/synthetic distinction may not be a psychological question, there is at least one aspect of this distinction that is psychological and of interest to semantic memory. Granting that formal semantics provides a basis for classifying sentences as analytic or synthetic, we may ask whether there is a mental procedure that reliably picks out all those sentences and only those sentences that have been classified as analytically true. This would constitute a psychological distinction between analytic and synthetic statements. But even given that such a procedure exists, the question arises of

whether it is a critical difference between set and network models. That is, only if all net models maintain a psychological distinction between analytic and synthetic statements and all set models blur this distinction would there then be support for the Glass and Holyoak proposal that the set-net distinction is chiefly about analyticity. However, an examination of existent models indicates that the set-net distinction is not correlated with this psychological-analyticity issue.

First, all semantic memory models, to our knowledge, have been applied to both analytic and synthetic statements. As we have already noted, net models, like the Marker Search theory, are intended as explanations of the way we verify statements like Some women are writers, which are purely synthetic, as well as analytically true statements like Some bachelors are unmarried. This is also true of the set theories proposed by Meyer (1970) and Smith, Shoben, and Rips (1974). This aspect of semantic-memory models is a reflection of the fact that the distinction between analytic and synthetic statements is not equivalent to one between propositions considered part of semantic memory and those thought to be a part of episodic memory (Smith, Rips & Shoben, 1974). Second, one may go on to ask whether an analytic/synthetic distinction can be formulated within the framework of set or net models. This can certainly be done, and it seems to be no more difficult for one class of models than for the other. In the case of network models, analytic statements might be recoverable by restricting the relations in the network to those which are true solely by virtue of the meaning of the concepts they connect. Similarly, for set models, analytic statements are those that can be confirmed by means of the semantic elements of features that are definitionally true of the

associated terms. Exactly the same arguments can be applied to the relation between binary truth values and the two classes of models we are considering. Hence, neither analyticity nor binary truth can be used to distinguish between set and network models, as Glass and Holyoak have proposed. As a consequence, neither type of model can be construed as evidence pro or con particular linguistic theories of semantics (e.g., Lakoff, 1972; Katz & Bever, 1975) that take sides on issues concerning truth-value systems. These issues in philosophy and linguistics, while important in their own right, are not at this time helpful in distinguishing among rival psychological theories.

What the Set-Net Distinction is About

Computation vs. pre-storage models. What then are the critical differences that divide set and network models? To get a grip on this problem, let us take a look at two simple semantic memory models. Figure 1 presents the Attribute theory, a set model described by Meyer (1970), along with Collins and Quillian's (1969) Hierarchical theory, a typical network model. Both were intended to account for the data obtained in a verification task. In such a task, subjects must decide on the truth or falsity of simple statements of the form An S is a P (where S designates a subject noun and P a predicate noun), and the data of interest are the reaction times and error rates. The Attribute model confirms a statement like A robin is a bird by comparing the features of the predicate category to those of the subject category, while in the Hierarchical model one verifies this statement by finding an acceptable path that links subject and predicate categories.

Insert Figure 1 about here

Note that in addition to their obvious representational differences, the two models differ in a rather striking respect. In the network model the proposition that a robin is a bird is represented directly in memory, and confirmation of the sample statement involves finding the corresponding proposition in memory. That is, the subset relation between robin and bird is not represented directly, and consequently it must be computed during the verification process. Since the differences just noted appear to hold for all set and network models, it seems that a critical difference between the two classes of theories is this: Network models posit that verification of subset relations can occur by searching for pre-sorted propositions, while set models assume that verification requires the computation of that relation.

We now need to specify a couple of boundary conditions on this Computation/Pre-storage dichotomy. First, no current network model of semantic memory assumes that all verifiable statements are confirmed by finding the corresponding proposition stored in memory. For such a position would imply that if someone can verify that Julius Caesar was a living thing, he must have at some time stored that exact proposition in memory. To avoid this claim, network modelers allow some room for computations. They posit inference routines that, when given stored propositions like Julius Caesar was a person and A person is alive, use the transitivity of subset relations to infer that Julius Caesar was a living thing. Thus in the network model in Figure 1, while the statements A robin is a bird or A bird is an animal would

be confirmed by searching for the pre-stored propositions, confirmation of A robin is an animal would involve an additional inference. Hence, all semantic memory models involve some computations. But we will continue to hold to our Computation/Pre-storage distinction since all network models assume that at least some subset statements are stored as single units in memory.

A second boundary condition concerns Computation models. In such models, not all relations are computable, for some meaning components must be pre-stored if the model is to compute anything. As an example, in the Attribute model, the features are pre-stored with their respective categories; these features can then be used to compute other relations, like the subset one.

Related distinctions. There are other distinctions that are correlated with our Computation/Pre-storage dichotomy. From our description of the models in Figure 1, it seems that the notion of a computation procedure leads to two consequences. First, since one cannot operate on the terms robin and bird directly, one must initially expand these terms into components that can be operated on (Rips, Smith & Shoben, 1975). In the Attribute model, the terms are expanded into sets of semantic features before any subsequent processing is done. The computation models of Schaeffer and Wallace (1970) and Smith, Shoben, and Rips (1974) also assume an initial expansion into semantic features, while some of the Computation models considered by Meyer (1970) assume that subject and predicate terms are first expanded into a list of exemplars of these terms, or else into the names of other items that share exemplars with the subject and predicate terms. In any event, all Computation theories assume some sort

of semantic expansion of the terms presented, and this is in contrast to most Pre-storage models.

The second consequence of positing a Computation procedure is that comparison processes are given a major role in verification (Rips, Smith & Shoben, 1975). In the Attribute model, once the subject and predicate terms have been expanded into sets of semantic features, these two sets must be compared to confirm that a subset relation holds between the two concepts. The notion of comparison processes is central to all Computation theories, and most of them further assume that variations in comparison processes are responsible for many of the empirical effects obtained in experiments on verification. While Pre-storage models also require comparison processes (so that the relations in the retrieved proposition can be checked against those in the test sentence), such processes play little role in the explanation of most empirical findings. Rather, variations in search processes are thought to underlie most findings of interest.

A third factor that correlates with the Computation/Pre-storage dichotomy has arisen as simple semantic-memory models, like those of Figure 1, have been revised to incorporate recent experimental results. For example, Rosch (1973) and Smith, Shoben, and Rips (1974) have found that the speed with which true sentences can be confirmed depends on how typical the subject category is of the predicate category. Thus, if apple is judged a more typical fruit than strawberry, An apple is a fruit should take less time to verify than A strawberry is a fruit. To cope with these results, network models have been broadened to allow pathways to be differentially accessible, where accessibility is determined by the co-occurrence frequency of the connected terms

(Collins & Loftus, 1975; Glass & Holyoak). Set models have also been revised by allowing the semantic features of a term to include those which characterize the concept as well as those that strictly define it (Smith, Shoben & Rips, 1974). Typicality effects are then explained on the basis of shared characteristic features between subject and predicate concepts. Thus, in explaining these typicality effects, a Computation model emphasizes a structural aspect, featural similarity, while a Pre-storage model stresses a functional aspect, co-occurrence frequency. Although it may be possible for Pre-storage models to incorporate a more structural account (see, e.g., Norman & Rumelhart, 1975), most of these models attribute typicality effects to co-occurrence frequency (Anderson & Bower, 1973; and the Marker Search model of Glass and Holyoak).

In summary, we have proposed four distinctions. For two of these--the Computation/Pre-storage contrast and the distinction based on semantic expansion--we know of little relevant data. As for the relative emphasis on comparison vs. search processes, this is a difficult issue to address directly, but it is related to Glass and Holyoak's recent experiments on disconfirming false statements. We will consider the relevant data in the section entitled "Experimental Studies of Disconfirmations" below. Lastly, we raised the issue of featural similarity vs. co-occurrence frequency as a means of explaining typicality effects. Here there are clearly pertinent data, and they will be discussed in "Criticisms of the Marker Search and Feature Comparison Models" below.

Characterizations of the feature comparison and marker search models. We now want to describe the Feature Comparison and Marker Search models in detail and show how they may be characterized by the above distinctions. Let us start with the Feature Comparison model. Its representational assumptions are quite simple. Each lexical term carries with it a set of semantic features. These vary continuously in the degree to which they confer category membership, with features at one extreme being essential for defining the concept, and features at the other extreme being only characteristic of the concept. Thus the term bird would include as defining features the notions that it is animate and feathered, and as characteristic features the notions that birds are of a particular size and have certain predatory relations to other animals (Rips et al , 1973; Smith, Shoben & Rips, 1974). More relevant to our proposed distinctions are the processing assumptions of the model. It is assumed that performance in a verification task is based upon a two-stage process. The first stage compares all of the features of the subject and predicate nouns in the test sentence, and assesses the degree of featural similarity between the two terms. In this stage, no consideration is given to whether the similar features are defining or only characteristic. It is next assumed that if the featural similarity is either very high (as in robin and bird) or very low (as in pencil and bird), then one can decide immediately whether a subset relation exists between the two nouns. That is, subject-predicate noun pairs with sufficiently high or low degrees of featural similarity will be classified as true or false, respectively, without going on to a second stage of processing. However, a second stage will be necessary for subject-predicate pairs that have an intermediate level of similarity

(as in penguin and bird, or bat and bird). The second stage considers only the more defining features, and determines whether all of the defining features of the predicate term match those of the subject term. This stage is thus identical to the simple Attribute model.

Clearly this model is a Computational one. The model essentially proposes that people have two ways of computing subset relations, where these two ways correspond to the two stages. Decisions based on only the first stage involve a heuristic computation, for such computations are rapid but may sometimes be in error (i.e., when many of the similar features are characteristic rather than defining). Decisions based on the second stage involve an algorithmic computation, for such computations are slow but consider only logically sufficient conditions. Both types of computation--heuristic and algorithmic--are alike, however, in that they require expansion of the lexical terms into underlying semantic features, and subsequent comparisons of these feature sets. The two types of computation differ in that the heuristic computation deals with characteristic as well as defining features. And it is these characteristic features that allow the model to explain typicality effects. That is, given that robin is judged to be a typical bird and chicken an atypical one, robin will presumably share more of the characteristic features of bird than will chicken. This will permit one to confirm A robin is a bird by means of only the heuristic process whereas the confirmation of A chicken is a bird will also require the time-consuming algorithmic computation. In sum, with regard to our distinctions, the Feature Comparison model has all the aspects of a Computation model, and these distinctions serve to elucidate certain of its key aspects.

As for the Marker Search model, its representational assumptions are more complex than any we have considered thus far. In the present model, each lexical term is represented by markers, a notion borrowed from Katz's (1972) theory of semantics. While Glass and Holyoak suggest that markers can be thought of as properties, in their own examples common words are directly associated with only a single marker. Thus the terms bird, chicken, and robin, are represented by the defining markers <avian>, <chicken>, and <robin>, respectively, where, for example, the marker <robin> would be characterized as "possessing the essential properties of a robin." A second representational assumption is that markers are interrelated so that one marker dominates or implies a set of other markers. As an example, <robin> implies <avian> which in turn implies <animate>, where the latter is the marker for animal. This implicational structure, which is intended to capture Katz's (1972) idea of redundancy rules, is ullustrated in Figure 2. There it

---

Insert Figure 2 about here

---

can be seen that the upshot of these assumptions is a semantic network similar to that of the Hierarchical model. However, further assumptions serve to distinguish the present theory from the Hierarchical one. The third representational assumption of the Marker Search model is that the hierarchical connections may sometimes be shortcut by direct pathways between nonadjacent markers. This is exemplified in Figure 2 by the shortcut path between <chicken> and <animate>. The final representational assumption is that information about contradictions is represented directly in the semantic

network. Specifically, a contradiction arises whenever two paths with the same label meet at the same marker, e.g., in Figure 2 <chicken> and <robin> contradict at <avian>.

The processing assumptions of the model are based on the notion that performance in a verification task is determined by a search of the semantic network. When a statement of the form An S is a P is presented, the subject accesses the defining markers of the two nouns and all other markers they imply or are implied by. In essence, this specifies a target section of the semantic network. This section is then searched, and the subject responds True as soon as he finds an acceptable path between the markers of the subject and predicate terms. Hence the time needed to confirm a true statement depends on the time it takes to find an acceptable path. This is just as it was in the Hierarchical model. However, unlike the Hierarchical model, if the shortcut path between <chicken> and <animate> is searched before the path between <chicken> and <avian>, the subject should be relatively quick in confirming A chicken is an animal, but relatively slow in confirming A chicken is a bird. Shortcut paths, then, provide a means of accounting for typicality effects. In a similar fashion, the subject responds False as soon as he finds a contradictory path between either (a) the defining markers of the subject and predicate terms (as in A robin is a chicken--see Figure 2), or (b) the defining marker of the predicate and a marker which implies the defining marker of the subject (as in A bird is a robin, where <chicken> both implies <avian> and contradicts <robin>--see Figure 2).

The above model is basically of the Pre-storage variety, as many propositions are represented directly in the network. Little expansion of

terms is needed for verification; rather, verification is a matter of searching for direct or indirect connections, or of searching for two connections that contradict one another. In all of these cases the critical determinants of verification times are the number of links in the pathways between markers and the order in which these pathways are searched. Thus, typicality and related effects can be explained in terms of the order in which certain shortcut paths are searched. That is, the probability that a particular shortcut exists, as well as its priority in the search order, increases with the co-occurrence frequency of the terms involved. Hence this theory differs from the Feature comparison model with respect to all of our proposed distinctions. The two models, then, should lead to different empirical consequences, and the next two sections of this paper are largely concerned with a comparison of the models with respect to certain empirical findings.

Criticisms of the Marker Search and Feature Comparison Models

The Glass and Holyoak paper contains (a) a detailed critique of the Feature Comparison model, and (b) a presentation of their own Marker Search model. In this section, we will first point out two potentially serious problems with the Marker Search model, and then attempt to rebut some of the criticisms of our own theory.

A Criticism of the Marker Search Model

In essence, the Marker Search model accounts for the existant data on disconfirmations by its notion of a contradiction, and for the data on confirmations by its ideas about the role of co-occurrence frequency in

determining short-cut paths and search order. We think both of these notions have their difficulties, as detailed below.

Contradictions: The encoding of negative information. The most important contribution of the Marker Search model is the way it handles false sentences, traditionally a problem for Pre-storage theories (see, e.g., Collins & Quillian, 1972; Anderson & Bower, 1973, chap. 12). As we have noted, the Marker Search model disconfirms statements by searching for tags on pathways that indicate two or more markers are contradictory. Although Glass and Holyoak have been hesitant to say exactly when two markers are contradictory, the only reasonable assumption seems to be that contradictory tags indicate which subsets of a common superordinate are disjoint (see Collins & Loftus, 1975). For example, the identically labeled paths from <chicken> and <robin> that intersect at <avian> in Figure 2 indicate that chickens and robins are disjoint subsets of birds. To see how this contradiction mechanism works in detail, it is convenient first to translate the language of Glass and Holyoak into more standard terminology. Accordingly, there are two ways of disconfirming statements in the model, one for sentences in which the subject and predicate categories are disjoint (e.g., All robins are chickens), and another for sentences where either the subject category is a superior of the predicate (e.g., All birds are robins) or the subject category partially overlaps the predicate one (e.g., All birds are pets). Disjoint statements are disconfirmed by searching for identically labeled links to a superordinate shared by the subject and predicate. For example, in disconfirming All robins are chickens, the subject locates paths from <robin> to <avian> and from

<chicken> to <avian> that have the same tag (see Figure 2). In contrast, superset or overlap statements are disconfirmed by searching for a subset category that is itself disjoint with the predicate category. For example, in disconfirming All birds are robins, a person must locate a subset of <bird> (e.g., <chicken>), and then determine that this subset is disjoint with <robin>, just as in the previous example.

While such a Pre-storage model for false sentences is a clear advance on earlier proposals, it is still possible to ask whether it is complete in the sense of being able to disconfirm all those sentences that we know to be false on semantic grounds. A consideration of some specific cases suggests it is not, and the simplest such example is illustrated in Figure 3a. Here we have four subsets (A, B, C, and D) of a single superordinate, S, such that A and B partially overlap, as do C and D. We indicate these set relations in

---

Insert Figure 3 about here

Figure 3a by a Venn diagram superimposed on the network structure. Given such a structure we can begin to label the paths, following the procedure that mutually exclusive subordinates of the same superordinate have the same labels. Since A and B partially overlap, they must have differently labeled paths to their superordinates, for if the tags were identical we would have evidence that A and B were disjoint. We indicate the overlapping status of A and B by placing α on the A-S path and β on the B-S path. Now however, we must decide how to label C-S and D-S. Using the rule that disjoint categories are indicated by the same tag, C-S must be labeled α since A and C

are disjoint by hypothesis. But if so, C-S and B-S will now have different labels, which indicates that B and C are not disjoint, according to our labeling procedure. This, however, contradicts our original assumption about the set relation between categories.

Clearly something is wrong with the original labeling rule, and we must consider other alternatives. One way out for the Marker Search model is to define away such a situation. For example, the model might posit that, for any overlapping categories (e.g., A and B), a new superordinate node, S', is formed together with the connections A-S', B-S', and S'-S, and that connections between A or B and S are disallowed. The resulting structure is illustrated in Figure 3b, labeled in a way consistent with our procedure. However, there are two major disadvantages to this modification. First, it posits memory nodes for no other reason than to bail out the model. We would need some evidence that such nodes actually represent concepts that play some substantive role in semantic memory. Second, the proposed modification prohibits the use of shortcut pathways in such situations. But we have seen that these shortcuts are warranted on other grounds, and are in fact a major structural assumption of the model.

However, there is a second possible way out of the present difficulties that we can explore. Suppose we allow multiple labels on a single path, so that C-S can be tagged by both α and β. If we assume that paths sharing at least one tag indicate disjoint subsets, then the structure in Figure 3c correctly reflects the relationships among A, B, and C. But we still have the C-S path to consider. If we label it with α or β, in order to show that D is disjoint with A or B, then D-S will also share labels with C-S. But

this means that D and C are disjoint sets according to our rule, and this contradicts the original hypothesis that C and D partially overlap. Our second way out has therefore led to only deeper difficulties, and so we have come up with no way in which the Marker Search model can provide an a priori basis for deciding when two paths have the same label.

In the course of our preceding arguments, we noted that the Marker Search model's provision for shortcut paths may, under certain assumptions, conflict with the method used to store negative information. A second way in which this conflict may arise is depicted in Figure 4, using an example

---

Insert Figure 4 about here

---

along the lines of Figure 2. In this diagram we have indicated the shortcut pathways between the nodes <canary> and <animate> and between <chicken> and <animate> by dotted lines. What is crucial here is the labeling of the paths terminating at <animate>. To indicate that <chicken> and <canary> denote disjoint subsets of animals, we have given both shortcut paths the label α. It follows that the <avian>-<animate> path must possess a different label (here, β) since neither <avian> and <canary> nor <avian> and <chicken> are disjoint subsets. But, then, what label should be used for the <mammalian>-<animate> path? The problem is similar to that raised with respect to Figure 3a. For if we use α in order to indicate that <mammalian> is disjoint from <chicken> and <canary>, we can no longer represent the fact that <avian> and <mammalian> are disjoint. Similarly, if we use β, we lose the ability to indicate that <canary> and <mammalian> and <chicken> and <mammalian> also

represent disjoint sets. Finally, as we have seen in the previous paragraph, using both α and β for the <mammalian>-<animate> path leads directly to further problems. It appears, therefore, that we must either prohibit disconfirmations on the basis of shortcut paths, or restrict or eliminate such paths entirely. Both possibilities violate the structural assumptions of the Marker Search model.

The problems associated with the structures in Figures 3 and 4 should not be taken to mean that it is impossible to store information about which subsets intersect and which are disjoint. Rather our demonstrations show only that the storage of negative information may not be as simple as markers on paths, as Glass and Holyoak's formalisms seem to suggest. It remains to be seen whether negative information can be incorporated into Pre-storage models in a way that is both theoretically parsimonious and consistent with experimental evidence. We note, by way of contrast, that such problems are not encountered by Computation models, since here the storage of negative information is unnecessary. Rather, negative decisions are made whenever defining features of predicate concepts mismatch those of subject concepts, as we have seen in terms of the Attribute and Feature Comparison models. We count this theoretical parsimony as a virtue of Computation models in general.[3]

The role of co-occurence frequency. As we have noted, co-occurence frequency plays a central role in the Marker Search model, as in other Pre-storage theories of semantic memory. Co-occurrence frequency determines what shortcut paths are formed as well as the order in which paths are searched, and these two factors, determine all of the empirical predictions from the model. That is, given co-occurrence frequencies, one should be able

to deduce the ordinal relations among reaction times for the verification of any set of true or false sentences. However, no norms of co-occurrence frequency have yet been published, and for this reason predictions from the Marker Search model have been generated from other, more readily available data. In particular, Glass and Holyoak rely on the frequency with which subjects produce a predicate noun when given a sentence frame containing the subject noun. For example, raters may be asked to complete the frame All birds are ? with a noun that will make the sentence true; the frequency with which a group of raters produce a particular predicate noun (e.g., animals) is then taken as an estimate of the co-occurrence frequency of the subject-predicate pair (e.g., of the birds-animals pair).

In an earlier paper (Smith, Rips & Shoben, 1974) we argued that co-occurrence frequency may not offer a satisfactory explanation of semantic phenomena because co-occurrence is itself determined in part by semantic factors. Thus, the words which appear in the present sentence co-occur because of the meaning relations they bear to one another and not because of the frequency with which they have been grouped. Frequency, therefore, may have the status of an epiphenomenon.

This anti-frequency argument is strengthened by reaction time effects with unfamiliar stimuli where co-occurrence frequency cannot be a factor. These effects must be attributed to structural aspects of the stimulus domain itself. Evidence on this score comes from a series of experiments by Rosch, Simpson, and Miller (1976), who used sets of dot patterns, stick figures, and letter strings as stimuli. To illustrate the critical findings, consider the case where letter strings were employed. Subjects first learned to classify

12 individual strings into two disjoint categories, and then were given a reaction time task in which they pressed one of two buttons depending on the category of a presented item. Finally, the subjects were asked either to rate the typicality of each of the instances, or to produce as many items as possible from each category. The strings themselves had been generated by varying the number of letters that a given string shared with other members of its category, and this variable (number of shared letters) determined all performance measures. Instances with more letters in common were learned in fewer trials, were classified faster, and had higher typicality ratings and production frequency than their counterparts. Similar results were obtained even when the less typical items were presented more frequently during initial learning. In this way, Rosch et al. reproduced the usual typicality effects varying only the internal properties of the stimulus domain, and this suggests that co-occurrence frequency may not be a necessary factor in determining typicality effects even in semantic-memory studies.

Co-occurrence frequency may not be a sufficient cause of typicality effects either, but to investigate this, we need a reliable index of co-occurrence frequency. The problem with the usual indices--production frequencies, as in Glass and Holyoak, or ratings of how often two terms seem to occur together, as in Anderson and Reder (1974)--is that they may be determined by semantic factors, as we noted earlier. There is, however, one index available that has the potential for providing an objective measure of co-occurrence frequency, the Kučera and Francis (1967) corpus of written American English (not to be confused with their simple word frequency counts). From this corpus we can tabulate the number of times an instance and its appropriate

category term appear together, which gives us a relatively direct measure of the kind of frequency we are interested in. In what follows, we will refer to this measure as the KF count.

The existence of the KF count allows us to assess certain claims about co-occurrence frequency and typicality effects. Suppose that: (1) co-occurrence frequency is indeed a sufficient cause of typicality effects, and (2) production frequencies and co-occurrence ratings are good estimates of objective co-occurrence frequency. Then it follows that: (3) the KF count should correlate with typicality effects, and (4) the KF count should correlate with production frequencies and co-occurrence ratings. Suppose instead that: (1') co-occurrence frequency is not a determinant of typicality effects, and (2') production frequencies and co-occurrence ratings primarily reflect semantic factors. Then it follows that: (3') the KF count should not correlate with typicality effects, and (4') the KF count should not correlate with either production frequencies or co-occurrence ratings, though the latter two indices should correlate with themselves as well as with typicality ratings.

To test these contrasting sets of predictions we used the data previously collected by Anderson and Reder (1974). These investigators collected reaction times (RTs) in a task where subjects were presented word pairs (e.g., turnip-vegetable), and had to decide whether the first item was a subset of the second. In addition to the RT data, Anderson and Reder also collected co-occurrence ratings ("how frequently do these two terms co-occur together?") and typicality ratings ("how typical is the instance of the category?"). This list of factors gives us everything we need to test our contrasting

predictions, except for production frequencies and KF counts. To obtain production frequencies, we used the norms collected by Battig and Montague (1969), who had subjects produce as many instances of given categories as they could in a 30 sec interval. Thirty-six of the 40 category terms used by Anderson and Reder correspond closely to categories in the Battig and Montague norms, and we will confine our subsequent analysis to these common categories.[4] Finally, we obtained our KF counts by defining an instance-category co-occurrence as the appearance of both terms within two lines of coded text (70 characters per line).

To test the contrasting sets of predictions, we simply carried out correlational analyses on the five factors mentioned: True RTs, typicality ratings, co-occurrence ratings, production frequencies, and KF counts. Consider our first set of predictions, where true co-occurrence (estimated by the KF count) supposedly determines typicality effects, as well as co-occurrence ratings and production frequencies. Contrary to predictions, the KF count did not correlate at all with True RTs, $\underline{r}(70) = .00$, and correlated only marginally with co-occurrence ratings, $\underline{r}(70) = .23$, and production frequencies, $\underline{r}(70) = .22$, $.05 < \underline{p} < .10$ in both cases. Thus the results offer little support for our first set of predictions, and are in far better agreement with our second set. Recall that in the latter, the KF count was not expected to correlate with RTs, co-occurrence ratings, or production frequencies, while all subject-generated measures were expected to be intercorrelated. In fact, all three subject-generated measures were substantially intercorrelated. Co-occurrence ratings correlated highly with production frequencies, $\underline{r}(70) = .66$ $\underline{p} < .01$, and with typicality ratings, $\underline{r}(70) = .70$,

$p < .01$, while production frequencies and typicality ratings were themselves intercorrelated, $r(70) = .63$, $p < .01$.

The above findings, then, favor our second set of predictions and the hypotheses that generated them: true co-occurrence frequency does not determine typicality, and subject-generated estimates of this factor reflect semantic factors. But there is reason to be cautious in drawing these conclusions. For our KF counts may be limited by the relatively small number of times our instance-category pairs actually appeared together in the Kučera and Francis corpus. However, there is an additional result in the literature suggesting that the KF count is not positively correlated with True RTs. This is the finding of Rosch et al. (1976) that for a completely different set of items, the KF count was negatively correlated with ratings of typicality; given this, and the fact that highly typical items are responded to quickly, it seems most unlikely that co-occurrence frequency is the cause of rapid responding to typical items. But still, until more work is done with the KF count, we shall have to settle for a cautious conclusion: There is no evidence that typicality effects are caused by co-occurrence frequency when this factor is measured by a relatively objective index.[5]

Even this weak conclusion leaves the Marker Search model (and all other Pre-storage models) without a theoretical explanation of the well-documented relations between RTs on the one hand, and typicality ratings and production frequencies on the other. This is in contrast to the Feature Comparison model, model, where featural similarity is assumed to be responsible for the effects of typicality ratings and production frequencies on True RTs. On this view, all of the subject-generated measures we discussed above are based on featural similarity, and that is why they are all correlated with True RTs, as well as

with one another. Furthermore, there are two pieces of evidence that directly link featural similarity to typicality ratings. First, Rips et al. (1973) showed that the features derived from a multidimensional scaling of a set of animal terms can predict typicality effects in semantic memory tasks (see also Shoben's subsequent scaling work, discussed in Smith, Rips, and Shoben, 1974). Second, there is the Rosch et al. (1976) study described earlier, where explicit variation in featural similarity induced concomitant variations in typicality ratings, as well as in production frequencies and RTs.

Criticisms of the Feature Comparison Model

In their paper, Glass and Holyoak refer to several sources of difficulty with the Feature Comparison model, apart from those problems associated with the Holyoak and Glass data. Some of these criticisms are concerned mainly with the evidence in support of the model presented in Smith, Shoben, and Rips, (1974). However, other remarks are addressed to the more general question of whether the Feature Comparison model is, in principle, able to account for verification of sentences other than subset statements. Both problems are obviously important ones, if they can be substantiated, and we deal with them in the following.

Can the Feature Comparison model be extended? According to Glass and Holyoak, the Feature Comparison model is inherently unable to encode relational information such as the notion of possession expressed by have in Elephants have ears. If so, the model could not explain how such statements are verified, and in addition, would have difficulty in accounting for the meanings of words that have relational components as part of their definition. But this supposed

limitation to non-relational components has never been part of the Feature Comparison model. Indeed, in an earlier paper (Smith, Rips & Shoben, 1974), we discussed sentences like the above example in some detail, as well as other sentence types commonly used in semantic memory (e.g., An ostrich is large). To rehearse our proposal concerning has, a predicate like has ears can be represented by an ordered pair, where the first member includes the semantic features of the verb (perhaps a single feature, has-as-a-part), and the second contains the feature list of the predicate noun. In verifying such a sentence, one would compare the features of the subject category to the representation of the predicate just described; if the subject category's features contain those of the compound predicate then the sentences will be true, and otherwise false. Thus, according to the model, sentences containing relational information can be encoded and, further, relational components can be part of the analysis of individual terms. In fact, in a new series of experiments, we have shown that the mechanics of the Feature Comparison model can be used to predict reaction times for the verification of sentences containing has (Rips, Shoben & Smith, 1975).

This, however, does not absolve the Feature Comparison model of all theoretical difficulties. It is merely that the problems faced by the model are not different in kind from those surrounding theories like the Marker Search model. As Glass and Holyoak acknowledge, these difficulties concern the way such models can be constrained so as to provide a principled account of semantic phenomena. For Pre-storage models, this comes down to specifying boundary conditions on permissible nodes and relations, as well as limits on the types of search procedures that can be employed. For Computation models, similar constraints must be established on the semantic components and

comparison processes. Thus the problem is not one of the generality of these models, but rather one of accounting for experimental data in other than an ad hoc fashion.

The empirical status of the Feature Comparison model. After a review of the relevant evidence, Glass and Holyoak conclude that there is little experimental evidence to support the processing assumptions of the Feature Comparison model. Their reasoning is as follows. The Feature comparison model identifies two factors that should, theoretically, influence RT; these include ratings of semantic relatedness, which should affect the first stage, and category size, which should affect the second stage. Neither factor, according to Glass and Holyoak, has been shown unambiguously to determine RTs, and therefore, no unambiguous evidence for the Feature Comparison model exists.[6]

These variables are important to the model, and a lack of evidence for them would indeed undermine the theory. Let us first consider the evidence for the effects of relatedness on semantic decisions. As Glass and Holyoak acknowledge, a large number of studies can be construed as showing effects of relatedness (e.g., Loftus, 1973; Meyer, 1970; Rips et al., 1973; Rips, Shoben & Smith, 1975; Smith, Shoben & Rips, 1974; Wilkins, 1971). But Glass and Holyoak argue that: (a) Ratings of semantic relatedness are sometimes less accurate predictors of RT than are production frequences (Smith, Shoben & Rips, 1974), suggesting that production frequency, not rated relatedness, is the key factor; and (b) Certain findings are more plausibly explained on the basis of search order than shared features (Glass, Holyoak & O'Dell, 1974; Loftus, 1973), again suggesting the importance of production frequencies

(supposedly measures of search order) over that of relatedness (supposedly a measure of shared features).

We have already considered the issue of relatedness vs. production frequency when we reanalyzed the results of Anderson and Reder (1974). There we found that the correlation of RT with relatedness (typicality) was actually slightly higher, though nonsignificantly so, than the correlation with production frequency (see Footnote 5). Previously, however, we have found one case where production frequency was a better predictor of RT than was semantic relatedness (Smith, Shoben & Rips, 1974, Experiment 1). So we have something of a discrepancy between these experiments with regard to whether a rating of relatedness or production frequency is the better predictor of RTs. This discrepancy may be due to any of a number of differences between the two experiments. However, even if production frequency was consistently superior to rated relatedness in predicting RTs, we believe that this would say little about the underlying mechanisms (search order vs. shared features) responsible for the RT effects. This is because production frequency norms are generally collected with subjects under speed pressure, just as they are in standard RT tasks. Consequently, extrinsic factors that affect all speeded tasks (e.g., factors that influence stimulus encoding) will increase the correlation between production frequency and RT. By contrast, subjects are usually not timed as they make relatedness judgments and are therefore uninfluenced by such extrinsic variables. For this reason, we might expect lower correlations between RT and relatedness than between RT and production frequency even if both ratings and frequencies were principally determined by shared semantic

features. In view of such considerations, the relative size of the correlations in question seems like an unimportant issue.

The second question concerning the role of semantic relatedness is whether this variable is sufficient to explain certain problematic findings. One set of findings (by Glass et al., 1974; and by Holyoak and Glass) exhibit cases in which reaction time decreases with relatedness for false sentences, a result that is contrary to the Feature Comparison model's predictions. We will discuss this evidence in the next section. The second kind of experimental evidence that seems counter to the Feature Comparison model is Loftus' (1973) demonstration of asymmetries between verifying that an instance is a category member and verifying that a category is the superordinate of an instance. For example, it is easier to verify that insect is a superordinate of the previously presented instance butterfly than to decide that butterfly is an instance of the previously presented superordinate insect. By contrast, it is easier to decide that shrimp is an instance of seafood (seafood presented first) than that seafood is the superordinate of shrimp (shrimp presented first). If RT is determined by relatedness, and if relatedness is itself a matter of shared features, why should such asymmetries arise?

There are, however, a number of ways to explain Loftus' result that are fully in keeping with the Feature Comparison model. First, we note that according to the original formulation of the model, the relatedness value computed in the first stage is based not on the number of shared features between instance and category, but on the proportion of the category's features that are shared (see Smith, Shoben & Rips, 1974). While this account was intended to apply to situations in which the instance and category were

presented simultaneously, it seems reasonable to suppose that when the items are presented in sequence, as in the Loftus experiment, relatedness should be determined by the proportion of shared features of whichever term is presented first. The two proportions need not be equal, of course, for they will depend on the total number of defining and characteristic features in the term presented first.[7]

A second explanation of Loftus' result is to assume that when the superordinate (e.g., insect) is presented first, subjects attempt to generate possible instances in anticipation of the to-be-presented instance. Similarly, when an instance (e.g., butterfly) is presented first, subjects generate possible superordinates. Whether subjects are successful in anticipating the correct item will depend on two factors: (a) the instance-superordinate relatedness, and (b) the number of alternative items with higher relatedness than the correct one. We can thus explain the asymmetry between butterfly-insect and insect-butterfly by appealing to the (b) factor. That is, there are more insect-instances with higher relatedness values than butterfly, than there are butterfly-superordinates with higher relatedness than insect. For the seafood-shrimp example, this ordering with respect to the (b) factor reverses. Again, instance-category asymmetries are not inconsistent with the Feature Comparison model.

The second factor questioned by Glass and Holyoak is category size. Category size predictions arise from the Feature Comparison model's second stage, where the defining features of the predicate are compared to those of the subject noun. The total number of the predicate's defining features should therefore determine second stage duration according to most serial and

parallel mechanisms. If we further assume, with Meyer (1970) and Clark (1970), that larger categories are likely to have fewer defining features than their subordinates, it follows that the duration of the second stage should decrease with increasing predicate size. For example, the time to complete the second stage should be greater for A bee is an insect than for A bee is an animal.

It is difficult, however, to test this prediction directly for two reasons. First, a simple change in the category size of the predicate is not sufficient, since such a change is likely to alter the subject-predicate relatedness and hence the probability that the second stage is even executed. Second, the second-stage difference that we are interested in may not occur on every trial; this is because some responses will always be made after only first-stage processing for there is as yet no experimental technique that ensures second-stage processing on every trial.

In view of these obstacles to a direct test of our category-size prediction, we attempted to assess it indirectly. In one attempt (Smith, Shoben & Rips, 1974, Experiment 1), we varied the size of the predicate categories in a standard verification task. Here, we used an analysis of covariance to eliminate any effects that category size might have had on relatedness. Contrary to predictions, we found no significant residual effect of category size when RTs were corrected in this way. In retrospect, this failure of the category size hypothesis seems surprising. The mathematical model presupposed by the analysis of covariance is not equivalent to that of the mathematical version of the Feature Comparison model itself, and so there is no guarantee that estimates of the category-size effects from the two mathematical procedures will coincide.[8] In order to derive estimates of category-size effects

from the mathematical version of the Feature Comparason model, we performed a second verification experiment and fit the model explicitly to the results (Smith, Shoben & Rips, 1974, Experiment 2). In this case the duration of the second stage for large categories (animal and plant) was calculated to be 161 msec, while the estimate for small categories (bird, insect, fruit, and vegetable) was 280 msec. So, as predicted, larger predicate categories were processed faster in the second stage. It should be noted that the model-fitting procedure itself did not constrain the former value to be smaller than the latter, so that these results constitute a confirmation of the underlying theory.

The parameter values just described were obtained by using error rates to help predict reaction times, following the procedure outlined by Atkinson and Joula (1974). This procedure has been criticized by Glass and Holyoak who claim that it trades on a general positive correlation between errors and RTs. However, several points can be made in response to this. First, recent evidence suggests that high positive correlations between errors and RTs are far from universal (Pachella, 1974). Second, even if this correlation were a truly general one, it is irrelevant in evaluating the crucial parameters of the model. Clearly, high correlations between errors and RTs imply nothing about the parameter values for the second stage that were discussed above. Finally, Smith, Shoben, and Rips also used a second procedure to predict the obtained data. In this procedure, error rates as well as RTs were predicted only from relatedness ratings. Here, there is no way we could have traded on a general positive correlation between errors and RTs, yet we still found that the estimated duration of the second stage was less for larger predicate

categories (245 msec) than for smaller ones (311 msec). Thus, there is in fact some evidence that the size of the predicate category affects semantic decisions. Such evidence fits nicely with the Computation models that assume semantic decisions are based on a comparison of features, with fewer features resulting in shorter comparison times. In contrast, it is not at all clear how Pre-storage theories like the Marker Search model would account for these results.

Experimental Studies of Disconfirmations

It remains for us to account for the empirical results of Holyoak and Glass on disconfirmation times, which, taken at face value, violate a major prediction of the Feature Comparison model. There are actually two sets of findings of interest, one concerning the disconfirmation of disjoint statements, the other concerning the disconfirmation of superset and overlap statements. We deal with each in turn.

Disconfirming Disjoint Statements

The Holyoak and Glass results. Using a standard verification paradigm, Holyoak and Glass presented subjects with 39 disjoint sentences of the form All S are P and 39 of the form Some S are P, in addition to other sentences that are irrelevant to the present issue. The 78 false sentences were subdivided by Holyoak and Glass into three types: high-production frequency, low-production frequency, and anomalous statements. These distinctions were based on an earlier experiment in which subjects were asked to generate completions for the sentence frames All S are ? and Some S are ? such that the resulting sentences were false. Holyoak and Glass then tabulated the production frequencies for these false completions. According to a

straightforward interpretation of the Marker Search theory, the frequency with which a particular completion is produced should reflect the amount of time necessary to disconfirm the corresponding sentence. For example, the production frequency of men to the frames All women are ? should predict the time needed to disconfirm All women are men, since completing frames involves finding a contradiction that is also used to disconfirm the statement in the verification task. Thus, high-frequency completions (produced by a mean of 35% of their 14 subjects) should be disconfirmed faster than low-frequency completions (produced by 5%), and these in turn should be falsified faster than anomalous completions (4%). Note, however, that the difference in production frequency between low and anomalous sentences is slight.[9]

Holyoak and Glass also obtained ratings of semantic relatedness for each of the disjoint subject-predicate pairs, and this allows us to generate rival predictions from the Feature Comparison model. These ratings show that the high-frequency sentences were somewhat more closely related than low-frequency sentences, and that low-frequency sentences were much more closely related than anomalous ones; the means were 4.88, 4.47, and 1.76, on a 7-point scale, for high, low, and anomalous sentences, respectively. Since the Feature Comparison model predicts that disconfirmation times should increase with relatedness, the high-frequency statements should take the longest to disconfirm, the low-frequency next longest, and the anomalous statements should be the fastest. This, of course, is the exact opposite of the ordering predicted by the marker Search model.

The results of this experiment disconfirmed major predictions of both models. First, contrary to the Feature Comparison model, low-frequency sentences took longer to disconfirm than the high-frequency ones. And second,

contrary to both theories, anomalous sentences took about the same amount of time as high-frequency sentences. Before commenting on an interpretation of these results, it seems important to inquire about their robustness. Since these findings are surprising ones, we decided to replicate them.

Experiment 1: A partial replication of Holyoak and Glass. We attempted to replicate the part of Holyoak and Glass's experiment that dealt with disjoint statements quantified by All, making only minor changes in procedure and design. We used a total of 132 word-pairs. Two sets of 39 pairs were selected from Holyoak and Glass. These sets comprised the disjoint statements and their true counterparts that were quantified by All in Holyoak and Glass. Our remaining 54 pairs were used as fillers to control for frequency of nouns in true vs. false items, such that: (1) all subject nouns were presented equally often in true and false items; and (2) approximately one-third of the predicate nouns appeared in a true item only, one-third in a false item only, and the remaining third once in both a true and false item. Only the 78 pairs selected from Holyoak and Glass were analyzed.

The two members of a pair were typed in uppercase Gothic in a single line (and were separated by a hyphen) on a 6" x 9" white index card. Subjects were instructed that a pair was to be considered True if the left-hand member was a subset of the right-hand one, and False otherwise. Twenty of the filler pairs were selected as practice items while the remaining 112 pairs were randomly ordered. The same order was used for all 20 subjects, who were Stanford undergraduates. The pairs were presented in a Gerbrands two-field tachistoscope at a viewing distance of 59 cm and each pair was preceded by a 1.5 sec fixation point. Responses were made on two telegraph keys, which,

when depressed, terminated the display of the pair and a Standard Electric Timer. The assignment of keys to response types (True and False) was balanced over subjects.

In analyzing the results, RTs to the critical 78 pairs were analyzed across both items and subjects (Clark, 1973), which separate analyses for True and False responses. The 39 true pairs were categorized as high-, medium-, and low-production frequency, following Holyoak and Glass's classification of these same items. This production frequency factor was a within-subjects variable, with stimulus pairs nested within frequency levels. In a similar vein, false pairs were divided into high, low, and anomalous items, again following Holyoak and Glass's classification.

Table 1 presentes mean True and False RTs from both the subjects and items analysis. The RTs differ slightly for our two analyses because we have

---

Insert Table 1 about here

---

used unweighted means analysis. For the False items, both sets of means show that RTs were fastest for anomalous pairs, next fastest for high-frequency pairs, and slowest for low-frequency items. The overall difference among the False means was significant at the .001 level (min $F'(2,59) = 8.09$), and Newman-Keuls analyses showed all pairwise comparisons among these means to be significant in both the analysis by items and that by subjects. This finding contradicts that of Holyoak and Glass, who found no significant differences between anomalous and high-frequency sentences.[10] Evidence for differences among the True means was more equivocal, as the min F' statistic

showed no significant differences (min $F'(2,59) = 1.86$, $p > .10$). However, the subjects analysis does indicate a difference among these means (see Table 1), and a Newman-Keuls analysis over subjects found high-frequency pairs to be faster than either medium- or low-frequency items. Further, a test for the linear trend among the True means showed a marginally significant effect (min $F'(1,56) = 3.71$, $.05 < p < .10$). The results for True RTs, then, are in rough agreement with Holyoak and Glass, who found the same ordering of means as we did.

The main discrepancy between Holyoak and Glass and our study concerns the relation between anomalous and high-frequency statements. It is possible that Holyoak and Glass failed to find a significant difference between these statement types because they repeated subject-predicate pairs in an unbalanced fashion. That is, considering both statements quantified by All and by Some in Holyoak and Glass, 9 (out of 17) high-frequency items, 6 (of 20) low-frequency items, and no anomalous items were repeated. If these repetitions decreased RT, high-frequency statements may have been artifactually fast relative to anomalous statements. In the present study, only statements quantified by All were used, and for these statements, there were no repetitions of subject-predicate pairs in the critical false statements.

Implications of Experiment 1. With respect to the two models of interest, our results can be summarized as follows. In congruence with the Feature Comaprison model, false items containing very unrelated nouns (the anomalous pairs) were verified extremely rapidly. This finding is contrary to the Marker Search model, which predicts relatively slow RTs to these pairs

because of their low production frequencies. However, our experiment duplicates Holyoak and Glass's important finding that low-frequency pairs are disconfirmed faster than high-frequency ones, despite the fact that high-frequency items are also rated as more related in Holyoak and Glass's norms. This finding is inconsistent with the Feature Comparison model, but in accord with the assumptions of the Marker Search model. Hence the results contradict some major predictions of both models. It seems that if we are to salvage the Marker Search model we must explain away the data from anomalous pairs. Alternatively, if we want to rescue the Feature Comparison model, we must explain the relation between high- and low-frequency pairs.

Let us first consider some ways to salvage the Marker Search model. Holyoak and Glass were aware of the problem that anomalous statements posed for their model, since this problem manifested itself in their own data (recall that they found anomalous statements were disconfirmed faster than low-frequency ones even though both statement types had comparable production frequencies). To reconcile these findings with their model, they proposed a new "admittedly ad hoc" device to the theory, namely that "...certain abstract types of information which differentiate between almost all words (such as the distinction between 'living' and 'non-living') are uniformly accessed quickly" (p. 237). Thus anomalous sentences should be disconfirmed quickly on the basis of this abstract information. But this leaves us with a serious question. If this abstract information is accessed rapidly, why do the production frequencies collected by Holyoak and Glass show anomalous completions to be rare? If production frequency is truly an indicator of search order, then anomalous completions should be fairly common, which is not the

case. Holyoak and Glass's reply is that "...production frequency may not be a valid measure of the association strength of such abstract properties" (p. 237), since these abstract concepts may not correspond to single lexical items in English and may be very rare in written or spoken language.

This explanation is not just ad hoc; it is almost surely incorrect. First, it is unclear why the supposed low frequency or lexical form of abstract predicates should result in low-production frequencies, since subjects contributing to these norms are not called upon to encode or produce these abstract terms at all. In order to produce an anomalous completion--chairs to the stimulus frame All birds are ? --the subject must determine that the superordinate pathways from <avian> and <chair> intersect at the appropriate abstract concept, like <thing>, with identically labeled paths; this is the only role played by the abstract concept. How quickly this can be done should depend on the order in which the abstract concept is searched, and this search order should not depend on factors like word frequency or number of words in the lexicalization of the concept. For if it did, the abstract concept involved should not be available quickly in verification tasks either, and this leaves Glass and Holyoak without any way of accounting for the falsification of anomalous statements. Second, abstract concepts need not be infrequent in English. Take, Holyoak and Glass' example of an anomalous statement, All birds are chairs. As we have just noted, disconfirming this sentence requires us to find a common superordinate where the pathways from the subject and predicate meet. Such a superordinate might be thing or object (not living thing, as Holyoak and Glass assume since living thing is the superordinate of the subject term only). Both thing and object have the

advantage of being single lexical items and fairly common, at least in written English (thing appears 333 times per million and object 65 time per million in the Kučera-Francis, 1967, norms). Note, in addition, that the abstract term thing will serve to disconfirm any anomalous sentence that is false by virtue of one concept being animate and the other inanimate, as in the above example. Third, the purpose of using production frequency as a predictor of RT was, according to Glass and Holyoak, to provide empirical constraints on their model. The importance of doing so, as they note, is that the Marker Search model lacks any structural constraints on search order. But Holyoak and Glass's abandonment of production frequency for abstract concepts seems to leave the model without empirical constraints of search order as well, allowing it to "predict" any RT results whatsoever. Thus the modification proposed by Glass and Holyoak to account for findings on anomalous statements is fraught with problems.

Now let us see what we can do to salvage the Feature Comparison model. Recall that its problem is that it cannot account for why high-frequency statements are disconfirmed faster than low-frequency ones. To get some leverage on this problem, let us consider in detail some of the Holyoak and Glass items. Low-frequency items included the sentences Some (All) women are babies, Some (All) valleys are lakes, Some (All) flowers are foods. In contrast to these difficult items, the high-frequency counterparts were Some (All) women are men, Some (All) valleys are mountains, and Some (All) flowers are trees. The subject and predicate concepts in both the high- and low-frequency sentences share a fairly large number of semantic dimensions. Women, men, and babies, for example, share those dimensions common to humans.

However, the subject-predicate pairs in high-frequency items (e.g., women-men) seem to possess directly opposing values on at least one shared dimension (sex, in our example). While it is possible to find such opposing values for the subjects and predicates of low-frequency pairs, the relationship is not as clear-cut. Thus, while women and babies may differ on the dimension of age, this difference depends on interpreting women in its most specific sense (woman as adult human female) rather than its more general one (woman as human female). This is the kind of intuition that led Glass and Holyoak to formulate the Marker Search model, and we concur that it is an important insight. The task for us is to determine some way of accommodating this intuition into the Feature Comparison model.

There are at least two ways of making this accommodation. The first is to change the model by adding some new content to the second stage. Specifically we may assume that this stage terminates as soon as any mismatching feature is found, and that a mismatching feature will be found sooner with high- than low-frequency statements. To use our previous example, the mismatching feature of sex may be found relatively quickly when comparing women and men, while the mismatching feature of age may be found relatively slowly when contrasting women and babies. Of course, added assumptions are of limited value unless they lead to new predictions, but the present assumption seems testable. It seems to predict faster confirmation times for the true statements Women are female and Men are male, than for Women are adults and Babies are nonadults. This seems like a reasonable prediction. But, alas, there are other problems with this approach. In addition to our assumptions

about self-termination, we must further assume that the faster second-stage processing of high- than low-frequency statements more than compensates for the greater likelihood of having to execute the second for high-frequency items (recall they have higher related values). While the small difference in relatedness between the high- and low-frequency completions argues for the viability of this approach, it is not an altogether satisfactory one without an explicit quantitative model.

Alternatively, we can accommodate the findings on high- and low-frequency pairs by altering our conception of first-stage processing, particularly of how one computes a relatedness value. We may assume that, in our Experiment 1, for example, when a subject computed the relatedness value, he gave more weight to dimensions with widely discrepant values than to dimensions with similar values. In this way, pairs like women-men would have been computed as less related than women-babies because of the extra weight given to the dimension of sex which differentiates the first pair. This method of computing relatedness may differ from that used by those subjects who contributed to the ratings, and were asked to rate "how closely you feel that two words are associated in meaning" (Holyoak and Glass's method, and our own). For in the latter situation, subjects may be inclined to give equal weight to all shared dimensions. This same ambiguity with respect to relatedness judgments has been noted by Fillenbaum (1973) in connection with multidimensional scaling techniques. To borrow Fillenbaum's example, the judged similarity of antonymous pairs like hot and cold will depend heavily on whether subjects attend more to the dimensions having similar values or to those having different values.

Again, we would like our proposed modification of the Feature Comparison model to lead to some new prediction(s). One such prediction is the following. If new rating instructions can be devised that induce subjects to emphasize dimensions having widely discrepant values, then this set of ratings should accurately predict the ordering of False RTs in Experiment 1. We attempted to test this prediction in Experiment 2.

Experiment 2: An alternative procedure for measuring relatedness. We asked 29 Stanford undergraduates to rate the relatedness of the subject-predicate pairs used in Holyoak and Glass's disjoint statements, by determining "...how easy it would be for the subject term to become the predicate." Our presumption was that such instructions would emphasize the importance of shared dimensions with discrepant values. The 63 distinct subject-predicate pairs were presented to the raters in a randomized list, and the raters were asked to produce a rating on a 10-point scale for each pair, with low values denoting more related items.

The results of this experiment may be summarized easily. The low-frequency pairs were now judged to be the most related, the high-frequency items next most related, and the anomalous statements least related of all. These relatedness ratings, then, display the same ordering as the RTs of Experiment 1, with relatedness being directly related to disconfirmation times as predicted by the Feature Comparison model. This was the case for both the set of items quantified by Some (used in the Holyoak and Glass study) and for the set quantified by All (used by Holyoak and Glass and by us in Experiment 1). For the former set, mean ratings were 4.20 for low-frequency

items, 4.54 for high frequency, and 7.71 for anomalous pairs (remember--low numbers mean high relatedness). For the latter set of items, means were 3.03 for low frequency, 4.43 for high frequency, and 7.11 for anomalous pairs. Differences between these means were significant in both cases, with $F'(2,53) = 30.08$, $p < .01$ for Some items, and $F'(2,48) = 27.20$, $p < .01$ for All items. Newman-Keuls tests showed that each of the pairwise differences within the two sets were significant, except for that between the high- and low-frequency pairs for statements quantified by Some.

So the Feature Comparison model is consistent with the pattern of means obtained in Experiment 1 as long as we assume that the relatedness value computed in the first stage mirrors the relatedness judgments provided by our subjects in Experiment 2. The relatedness norms collected by Holyoak and Glass fail to predict the results of Experiment 1 because their ratings reflected only the overall proportion of shared dimensions. Ratings of this kind have proved successful in earlier studies (e.g., Rips et al., 1973; Smith, Shoben & Rips, 1974), possibly because the false items used in the earlier studies did not discriminate between the two sorts of relatedness. These conclusions, however, need further scrutiny. Introducing yet another measure of the semantic relation between subject and predicate nouns may raise as many problems as it solves. Many types of ratings have been found to correlate with RTs for semantic memory judgments (e.g., co-occurrence ratings, production frequencies, relatedness ratings), and all of these measures are intercorrelated. So additional experiments and analyses are needed to tease apart the critical differences between these measures, and additional thought must be given to the factors responsible for the differences. Perhaps such

studies will show us that different measures all tap different, but equally important, semantic aspects, for there is surely no reason to think that there is one best measure of semantic processing.

Disconfirming Superset and Overlap Statements

The Holyoak and Glass results. There is one last set of findings due to Holyoak and Glass that we must still deal with. These findings involve RTs to false superset statements, e.g., All women are mothers, and false overlap statements, e.g., All women are writers, (what Holyoak and Glass call Counterexample statements). As we noted earlier, the Marker Search model assumes that these types of sentences are disconfirmed by a search for a subset of the subject category that is disjoint with the predicate category, as indicated by identically labeled pathways. The Feature Comparison model would disconfirm these kinds of statements in the same way it falsifies disjoint statements, that is, by finding mismatching features. Again the two models differ in the predictions they make about the disconfirmations of interest. But to see this, we need to examine the Holyoak and Glass study in detail.

In their study of how superset and overlap statements were disconfirmed, Holyoak and Glass's experimental strategy again involved finding cases where production frequency and relatedness ratings make discrepant predictions for False RTs. But in this case, Holyoak and Glass's frequency measure of interest is obtained by somewhat indirect means. Instead of using the frequency of completions that make a sentence frame false (the procedure used for the disjoint statements previously discussed), Holyoak and Glass use the frequency

of completions that make the corresponding Some statement true. (Note that a superset or overlap statement is true when quantified by Some; it is only false when quantified by All). For example, to predict RTs for disconfirmations of sentences like All fruits are oranges, Holyoak and Glass used completions of the frame Some fruits are ? that serve to make the sentence true. In theory, the higher the frequency of the Some completion, the faster one can disconfirm the corresponding false All statement. If, for example, apples is a high frequency completion to Some fruits are ? , then apples is readily accessible from fruits, and rejection of All fruits are oranges should be fast. This follows from the Marker Search model's hypothesis that such sentences are disconfirmed by a "back-up" search from the subject term (fruits) that finds a category (apples) that is disjoint with the predicate term (oranges).

The results showed significant effects on False RT of these true-completion frequencies. However, the False RTs showed no significant effects of production frequency of the false sentences themselves. For example, reaction time to disconfirm a sentence like All fruits are oranges was unaffected by the frequency with which subjects generate oranges to the frame All fruits are ? when asked to make the sentence false. This contrasts with the finding obtained for disjoint statements. Also, according to the norms collected by Holyoak and Glass, ratings of semantic relatedness coincided with the frequency of false completions, so they likewise failed to predict the False RTs. Holyoak and Glass note, however, that a significant residual effect of relatedness remains when the true-completion frequencies are controlled.

Holyoak and Glass conclude that these findings support the Marker Search model, because the model correctly predicts that False RTs should decrease with true-completion frequencies. They further conclude that the results conflict with the Feature Comparison model, because the latter cannot account for the effects of true-completion frequencies. We disagree, as we think the findings pose difficulties for both models. Consider first the Marker Search model. The problem here is how to explain the lack of effect of false-completion frequency. Surely, according to this model, false completions to All fruits are ? indicate the order in which pathways are searched from fruit, as this search must locate not only a potential predicate, like orange, but also another subset of fruit, like apple, that is disjoint with the potential predicate. For if this last step were omitted, it would be impossible to determine that the potential predicate actually made the sentence false. Thus the search needed to produce a false superset or overlap completion mirrors the search necessary to disconfirm the completed sentence. By the usual Marker Search logic, this should mean that these completion frequencies will predict RTs for the corresponding full sentences, and this is contrary to the obtained findings. Although Holyoak and Glass seem to dismiss this implication from their model, we think it provides an important disconfirmation of their theory.

Now consider the Feature Comparison model. The problem here is how to account for the effect of true-completion frequency on False RTs, since Holyoak and Glass were able to demonstrate that this effect was independent of rated relatedness. We could proceed as we did before, and attempt to add

some new assumptions to our theory so as to make it consistent with the problematic effect. In this case, however, we think this approach is unnecessary at this point in time. For we dispute the very claim that the effect of true-completion frequency has been adequately demonstrated, because there are two methodological problems with this study that undermine its principal finding.

First, it appears that several of the items in the critical set were simply misclassified by Holyoak and Glass. (All of the items are listed in their Appendix.) For example, the sentence All fruits are citrus is listed as one for which there was no high-frequency true-completion to Some fruits are ? that was disjoint with citrus. However, apple, listed as a high-frequency true-completion, seems to fill the role of such a disjoint predicate. A similar problem applies to All birds are swimmers, which was also classified as having no high-frequency true-completion associated with it; here robins seems to be such a completion. Removing just these two items from the critical sentences reduces the overall difference from 109 to 76 msec. This reduced effect is not significant over either subjects ($F(1,13) = 3.41$, $p < .05$) or items ($F(1,20) = 2.51$, $p < .10$). However, the effect is still in the right direction, and it might prove significant in future experiments that involve more subjects and more items. So while this problem is somewhat serious, it may not be that severe.

The second methodological problem is more bothersome. Holyoak and Glass did not control for the type of set relation within their critical sentences (those where true-completion frequency and relatedness were unconfounded).

That is, for the critical sentences, 8 of the 12 items with high true-completion frequencies were superset statements while the remainder were overlaps; in contrast, only 3 of 12 items with low true-completions were superset statements while the rest were overlaps (see Holyoak & Glass, Footnote 4). Thus true-completion frequency was confounded with the prevalence of superset statements. If we assume that superset statements can be confirmed faster than their overlap counterparts, we have come up with an alternative explanation of the problematic result. Experiment 3 provides support for this assumption.

Experiment 3: Differences between superset and overlap statements. Thirteen subjects (Stanford undergraduates) were given 150 statements to verify, all of the form All S are P. Half the statements were true, and half false. The false items contained 25 disjoint, 25 superset, and 25 overlap statements, and these three statement-types were equated for average subject-predicate relatedness as determined by previously obtained ratings.[11] Also, the average relatedness of subject-predicate pairs in the false statements (6.5 on a scale of 1-10, where high numbers indicate similar meanings) was roughly equal to the average relatedness of subject-predicate pairs in true statements (6.9). Each statement was presented only once, and there were no repetitions of words across the 150 statements.

Each full statement was typed in uppercase Orator in a single line on a 6" x 9" white index card. Subjects were simply instructed to decide whether the statement was True or False. The same random ordering of the 150 statements was used with all subjects. The statements were presented

in an Iconix three-field tachistoscope at a viewing distance of 68 cm, and each item was preceded by a 1 sec fixation point. The response panel contained three telegraph keys arranged horizontally. The middle key was used by the subject to initiate each trial, while the left and right keys were used to indicate True and False responses. All subjects used the key corresponding to their dominant hand to indicate true decisions.

The False RTs are the only ones of interest, and they were analyzed across both items and subjects. For the subjects analysis, disjoint statements were disconfirmed fastest (1510 msec), superset statements next fastest (1575 msec), and overlap statements were slowest of all (1721 msec). The overall effect of set relation was significant at $p < .01$, with $F(2,48) = 9.67$. Furthermore, subsequent planned comparisons showed that overlap statements were significantly slower than superset statements, $F(1,48) = 14.49$, $p < .01$, while the superset and disjoint statement-types were not significantly different from one another, $F(1,48) = 2.87$, $p < .1$. For the items analysis, the mean RTs for disjoint, superset, and overlap statements were 1506, 1580, and 1680, respectively; the effect of set relation was marginally significant, $F(2,144) = 2.42$, $.05 < p < .10$. While a planned comparison did not reveal a significant difference between overlap and superset sentences, $F = 2.34$, $.10 < p < .20$, the difference between them is of course in the expected direction, and the magnitude of the difference (100 msec) is relatively substantial for this kind of experiment. Lastly, the error rate on overlap statements (24%) was far greater than that on superset statements (8%), $t(12) = 4.35$, $p < .001$.

All things considered, these results indicate that overlap statements are harder to process than their superset counterparts, and this provides an alternative explanation of the Holyoak and Glass results. Thus there is no clear-cut evidence in the Holyoak-Glass study for the effects of true-completion frequency, or for what they have called "disconfirmation by counterexample." As we see it, this reduces the credibility of their theoretical claims.

How do the present results line up with the Feature Comparison model? It seems that they remove one problem for the model--the need to explain the effect of true-completion frequency on False RTs--and create a new one--the need to explain the effects of set relation on False RTs. That is, there is nothing in the Feature Comparison model that would lead us to expect that False RT should increase from disjoint to superset to overlap statements, when all three statement-types are equated for relatedness. Before trying to add some new assumptions to our model to account for these new results, it is helpful to localize the effects of set relation within the processes of the model. Two aspects of Experiment 3 suggest that set relation affected only the second stage of the model. First, all three statement-types were equated for relatedness, and, in terms of the model, this means that all false statements were equally likely to require second-stage processing. Second, as previously noted, true and false statements had roughly the same level of subject-predicate relatedness, and, according to the model, this means that many of the True-False decisions must have been based on second-stage processing (Smith, Shoben & Rips, 1974; Smith, Rips & Shoben, 1974).

We also have additional evidence that argues for a second-stage locus of the effect of set relation. To appreciate this evidence, consider the consequences of changing the quantifier used in Experiment 3 from All to Some. We have argued elsewhere that Some statements probably require a different second stage than that used in verifying All statements (Smith, Rips & Shoben, 1974). Essentially, this is due to the fact that the second stage used with All statements establishes a subset relation, and Some statements are true even when they manifest only a superset or overlap relation. Consequently, if the set-relation effect is due to the second stage, then this effect might not obtain if the quantifier is switched from All to Some. Accordingly, we basically redid Experiment 3 using Some as the quantifier. (To insure that True and False responses were still equally probable, we used only 25 subset statements and increased the number of disjoint statements to 75.) The results were simple. There was no longer any effect at all of set relation. If we restrict our attention to the 25 disjoint, superset, and overlap statements that were previously used in Experiment 3, the new means are as follows: for the subject analysis, disjoint = 1482 msec, superset = 1494 msec, and overlap = 1499 msec, $F(2,48) < 1$; for the items analysis, disjoint = 1502 msec, superset = 1522 msec, and overlap = 1517 msec, $F(2,48) < 1$. These results, then, line up with the notion that the set-relation effect of Experiment 3 was due to the second stage.

To explain why second-stage processing is faster for disjoint and superset statements than for overlap ones, it seems we must assume this stage is self-terminating. Disjoint statements would then be disconfirmed rapidly

if we further assume that general features, like being animate or being alive, are compared first since most of our disjoint pairs differed on these features. (Note that this differs from the Holyoak and Glass assumption about general features being accessed first, since we hold that such features only become available after an extensive amount of processing--the first stage--has been completed.) There is some support for our assumptions in Shoben (1974), where disjoint noun-pairs were disconfirmed faster when they differed on general features rather than just specific ones. Shoben, though, did not establish that this effect was independent of relatedness, so our assumptions should be considered speculative until further research is done.

It is somewhat more difficult to come up with an explanation of why second-stage processing should be faster for superset than overlap noun-pairs, as both types of noun pairs contained virtually no mismatches on general features. There is, however, one notable difference between the sets of defining features for superset and overlap pairs. In superset pairs the predicate term should contain more features than the subject term (as the predicate term is in fact a subset of the subject), while this imbalance need not hold in overlap pairs. Detection of this imbalance would provide sufficient grounds for disconfirming a statement, for obviously all the features of the predicate cannot be found among those of the subject if there are more features in the predicate to begin with. Thus it is possible that superset statements were processed faster than overlap ones because the subjects of Experiment 3 were sensitive to this imbalance, and terminated their second-stage processing as soon as the imbalance was detected. This is

a very ad hoc assumption, and again further research will be needed to determine if it has any merit.

To summarize, Experiment 3 appears to undermine the results of Glass and Holyoak on true-completion frequencies (Counterexamples, in their terminology); it also leads to some new problems for the Feature Comparison model, problems that call for further embellishments of the proposed second stage.

Summary and Future Directions

We began with an attempt to classify semantic-memory models, and after considering various unsatisfactory classifications, we proposed a distinction between Computation and Pre-storage models. Computation models, of which the Feature Comparison model is a current exemplar, emphasize semantic expansion of terms during sentence verification, and account for RT effects in verification experiments by means of variations in the time needed for comparison operations between these expanded concepts. Obtained effects of relatedness or typicality are explained by similarities among the elements of the expanded concepts. Pre-storage models, such as the Marker Search model, explain RT effects in terms of variations in search procedures that operate on a database of stored (usually interconnected) propositions. Typicality and kindred phenomena are explained away by means of co-occurrence frequency.

What is the current status of these models, in light of the evidence reviewed here? Much of this evidence related indirectly to the question of whether RT effects are best ascribed to search or comparison processes. But though we were able to offer a detailed contrast between a theory emphasizing comparison processes (the Feature Comparison model) and one emphasizing search

processes (the Marker Search model), we are not able to settle the comparison vs. search issue in any final way. There is no decisive empirical result that infirms either theory. This is not so much because semantic-memory research has uncovered no interesting facts, but rather because the models have been retrenched in an effort to account for the new facts. As a result, one of the outstanding questions is whether the revised models are too general to be testable, a problem that seems to be particularly acute for Pre-storage models (Glass & Holyoak, and particularly Collins & Loftus, 1975). For these models, there are no structural constraints at all on search order. Even the empirical constraints proposed by Glass and Holyoak can be by-passed by invoking extra mechanisms like those needed to account for disconfirming anomalous sentences. For Computation models, one of the important remaining problems is to specify the mechanics of the comparison process through further discriminating experiments; hopefully the present Experiments 1-3 begin to do this.

With respect to an explanation for typicality effects, we seem to be on firmer ground. There is no evidence whatsoever for the role of co-occurrence frequency, at least when frequency is measured in an objective way (as in our reanalysis of Anderson & Reder, 1974, or in Rosch et al., 1976). Although dependence on co-occurrence frequency is a relatively peripheral feature of Pre-storage models, the lack of evidence for co-occurrence frequency leaves these models without a principled explanation for their own structural organization. For example, the Marker Search model is left without any theoretical underpinning for its short-cut pathways or search order,

beyond the sheer need to account for the data. By contrast, those Computation models that take featural similarity as their starting point, have little difficulty in coping with typicality effects and other related phenomena. Instead, the problems faced by the latter models have to do with specifying the status of the features themselves, and the boundary conditions on feature combinations and feature comparisons.

Of course the issues contended by Computation and Pre-storage models do not exhaust the range of questions concerning psychological semantics. Nor do we need to resolve the former before pursuing the latter. For example, little experimentation has been done on rules for combining propositions semantically in complex sentences, and to our knowledge, no semantic-memory model has even explicitly addressed this problem. While we have done some preliminary work in this area (Rips, Shoben & Smith, 1975), there is no way at present to evaluate semantic-memory models on this issue. Similarly, most current models of semantic memory have been content to divide sentences into property statements (e.g., Oranges are round) and class inclusion statements (e.g., Oranges are fruit). But among the so-called property statements are a wealth of distinct semantic types, including modals (Oranges can roll), sentences with relative adjectives (Oranges are small), and sentences with complex verbs (Oranges grow). We know from linguistic and philosophical analyses that such sentences contain important semantic characteristics, yet we have no evidence at all concerning psychological distinctions among them. It seems to us that semantic memory has nothing to lose by dealing with a broader range of phenomena.

References

Anderson, J. R., & Bower, G. H. Human associative memory. Washington, D.C.: Winston, 1973.

Anderson, J. R., & Reder, L. M. Negative judgments in and about semantic memory. Journal of Verbal Learning and Verbal Behavior, 1974, 13, 664-681.

Atkinson, R. C., & Juola, J. F. Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce and P. Suppes (Eds.), Contemporary developments in mathematical psychology. San Francisco: Freeman, 1974.

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecitcut Category Norms. Journal of Experimental Psychology Monograph, 1969, 80, (3, pt. 2).

Clark, H. H. Word associations and linguistic theory. In J. Lyons (Ed.), New horizons in linguistics. Baltimore, Md.: Penguin, 1970.

Clark, H. H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 335-359.

Collins, A. M., & Loftus, E. F. A spreading-activation theory of semantic processing. Psychological Review, 1975, 82, 407-428.

Collins, A. M., & Quillian, M. R. Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 240-248.

Collins, A. M., & Quillian, M. R. Experiments on semantic memory and language comprehension. In L. W. Gregg (Ed.), Cognition in learning and memory. New York: Wiley, 1972.

Fillenbaum, S. Problems in the use of similarity data for the study of semantic structures (Tech. Rep. 114). Chapel Hill, N.C.: University of North Carolina, L. L. Thurstone Psychometric Laboratory, February 1973.

Frege, G. On sense and reference. Zeitschrift für philosophie and philosophische kritik, 1892, 100, 25-50.

Glass, A. L., & Holyoak, K. J. Alternative conceptions of semantic memory. Cognition, 1975, 3, 313-339.

Glass, A. L., Holyoak, K. J., & O'Dell, C. Production frequency and the verification of quantified statements. Journal of Verbal Learning and Verbal Behavior, 1974, 13, 237-254.

Hollan, J. D. Features and semantic memory: Set-theoretic or network model? Psychological Review, 1975, 82, 154-155.

Holyoak, K. J., & Glass, A. L. The role of contradictions and counterexamples in the rejection of false sentences. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 215-239.

Katz, J. J. Semantic theory. New York: Harper & Row, 1972.

Katz, J. J., & Bever, T. G. The rise and fall of empiricism. In T. G. Bever, J. J. Katz, and D. T. Langendoen (Eds.), An integrated theory of linguistic ability. New York: Crowell, 1976.

Kučera, H., & Francis, W. N. Computational analysis of present-day American English. Providence: Brown Univeristy Press, 1967.

Lakoff, G. Hedges: A study in meaning criteria and the logic of fuzzy concepts. Papers from the Eighth Regional Meeting, Chicago Linguistic Society. Chicago: University of Chicago Linguistic Department, 1972.

Loftus, E. F. Category dominance, instance dominance and categorization time. Journal of Experimental Psychology, 1973, 97, 70-74.

Meyer, D. E. On the representation and retrieval of stored semantic information. Cognitive Psychology, 1970, 1, 242-299.

Norman, D. A., & Rumelhart, D. E. Explorations in cognition. San Francisco: Freeman, 1975.

Quine, W. V. O. Two dogmas of empiricism. In W. V. O. Quine, From a logical point of view. Cambridge, Mass.: Harvard University Press, 1953.

Pachella, R. G. The interpretation of reaction time in information-processing research. In B. H. Kantowitz (Ed.), Human information processing: Tutorials in performance and cognition. Hillsdale, N.J.: Erlbaum, 1974.

Rips, L. J., Shoben, E. J., & Smith, E. E. Semantic distance and the verification of semantic relations. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 1-20.

Rips, L. J., Shoben, E. J., & Smith, E. E. Performance models of semantic composition. In R. Grossman, J. San & T. Vance (Eds.), Papers from the parasession on functionalism. Chicago: Chicago Linguistic Society, 1975.

Rips, L. J., Smith, E. E., & Shoben, E. J. Set-theoretic and network models, reconsidered: A comment on Hollan's "Features and semantic memory." Psychological Review, 1975, 82, 156-157.

Rosch, E. On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), Cognitive development and acquisition of language. New York: Academic Press, 1973.

Rosch, E. H., Simpson, C., & Miller, R. S. Structural bases of typicality effect. Journal of Experimental Psychology: Human Perception and Performance, 1976, in press.

Rumelhart, D. E., Lindsay, P. H., & Norman, D. A. A process model for long-term memory. In E. Tulving & W. Donaldson (Eds.), Organization and memory. New York: Academic Press, 1972.

Schaeffer, B., & Wallace, R. The comparison of word meanings. Journal of Experimental Psychology, 1970, 86, 144-152.

Shoben, E. J. Semantic features in semantic memory. Unpublished doctoral dissertation, Stanford University, 1974.

Smith, E. E., Rips, L. J., & Shoben, E. J. Semantic memory and psychological semantics. In G. H. Bower (Ed.), The psychology of learning and motivation (Vol. 8). New York: Academic Press, 1974.

Smith, E. E., Shoben, E. J., & Rips, L. J. Structure and process in semantic memory: A featural model for semantic decisions. Psychological Review, 1974, 81, 214-241.

Sternberg, S. The discovery of processing stages: Extensions of Donder's method. Acta Psychologica, 1969, 30, 276-315.

Tarski, A. [The concept of truth in formalized languages.] (J. H. Woodger, trans.). In A. Tarski, Logic, semantics, metamathematics. Oxford: Clarendon, 1956.

Wilkins, A. T. Conjoint frequency, category size, and categorization time. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 382-385.

Footnotes

[1]The research reported in this paper was supported by U.S. Public Health Service Grant MH-19705 and a grant from the Spencer Foundation, and in part by the National Institute of Education under Contract No. US-NIE-C-400-76-0116. We thank John Anderson and Lynne Reder for making their data available to us.

[2]All further references to Glass and Holyoak are to this paper. Similarly, references to Holyoak and Glass refer to Holyoak and Glass (1975).

[3]The problems that we have just discussed all stem from the procedure of assigning mutually exclusive subsets the same marker. One could argue that, rather than relying on this kind of procedure, we need instead to use our intuitions to decide when two paths are contradictory. The difficulties with this solution are obvious: it forgoes any *a priori* determination of contradictory pairs, and it may not lead to many predictions if intuitions about contradictions are not clear-cut. Both of these problems could eventuate in a theoretical formulation that lacks testability.

[4]It is possible to maintain that other measures of production frequency would have been more appropriate. Our choice was dictated by the availability of the Battig-Montague entries for the items used by Anderson and Reder (1974) and by the role that these norms have played in previous studies of semantic memory (e.g., Wilkins, 1971).

[5]There are some additional results from our correlational analysis that deserve comment. First, both production frequency and typicality ratings correlated with True RTs, $r(70) = -.23$, $.05 < p < .10$, and $r(70) = -.25$, $p < .05$, respectively. Neither of these findings is the least bit novel (see Smith, Shoben & Rips, 1974), though both correlations are surprisingly low in

light of previous results. Second, there was also a marginal positive correlation between typicality ratings and the KF count, $r(70) = .21$, $.05 < p < .10$. This result conflicts with the negative correlation of Rosch et al. (1976) that was just mentioned in the text. But the conflict may be more apparent than real, since the instance-category pairs of Rosch et al. covered a wider typicality range than did those of Anderson and Reder. So the Rosch et al. results may be the more sensitive ones. Finally we should mention that we performed a step-wise multiple regression of RT on the entire set of independent variables discussed above, plus measures of simple word frequency (determined by the standard Kučera and Francis norms). Typicality ratings entered the regression equation first, and of the remaining variables only the word frequency of the instance term showed even a marginally significant correlation with RT, $r(69) = -.21$, $.05 < p < .10$. Thus, typicality ratings taken together with instance word frequency seem to provide the best account of the RT data, $R = .32$, $F(2,69) = 3.96$, $p < .05$.

[6]Glass and Holyoak argue this in terms of Sternberg's (1969) Additive Factors method. We feel that invoking the Additive Factors method here may be something of a red herring. In the Feature Comparison model, the output of the first stage (the semantic relatedness value) directly affects the duration of the second stage; this means that if relatedness is manipulated experimentally, its effect will interact with any factor (e.g., category size) that influences the second stage. The Feature Comparison model, therefore, cannot be faulted for lack of additive effects between relatedness and category size since none are predicted. Of course, it may be possible to find

some third variable that influences the duration of the first stage without changing the relatedness value, and if so, this factor should produce additive effects with second stage variables.

[7]We could equally well assume that relatedness depends on the proportion of the second item's features that are shared. This in no way affects the present argument.

[8]More specifically, the estimate of the category size effect derived from the analysis of covariance is obtained by fitting an equation of the following form to the reaction time data (ignoring error):

$$RT_{ij} = t_{mean} + t_{size_i} + b(X_{ij} - \bar{X}). \qquad (a)$$

RT_{ij} is the reaction time to verify sentence j with predicate category size i, and X_{ij} is the relatedness rating for the same sentence. Here, t_{mean} is the overall mean RT, $\bar{X}$ is the mean relatedness rating, and t_{size_i} is the effect of category size. By contrast, the model proposed by Smith, Shoben, and Rips (1974, Equations 3 and 6) is more complex, and predicts True reaction times as:

$$RT_{ij} = t_{mean} + t_{size_i}\left[\frac{\Phi(c_1 - b_k X_{ij} - a_k) - \Phi(c_o - b_k X_{ij} - a_k)}{1 - \Phi(c_o - b_k X_{ij} - a_k)}\right], \qquad (b)$$

where Φ represents the normal distribution function, and a_k, b_k, c_o, and c_1 are parameters of the model. A similar equation obtains for False reaction times. The relationship between reaction time, relatedness ratings, and the estimate of the category size effect is clearly different in the two models,

and consequently, there is no reason to suppose that the estimates of t_{size} will be equivalent.

[9]At one point, Glass and Holyoak argue that the results for anomalous statements are not critical to their model because they wish to restrict their theory to explanations of high- and low-frequency statements. We are not convinced that this restriction is a principled one, and so we will consider the results for anomalous statements.

[10]Interestingly, though these results fail to replicate the Holyoak and Glass results for anomalous statements, they do replicate an earlier study of Glass et al. (1974). These authors, using noun-property statements rather than noun-pairs, found that anomalous statements were disconfirmed faster than high-frequency statements, which in turn were faster than their low-frequency counterparts.

[11]To keep matters comparable to Holyoak and Glass, relatedness was determined by ratings of "closeness in meaning"--the standard procedure. As an afterthought, we also measured relatedness by the ratings used in Experiment 2; these ratings also showed that the superset and overlap statements were equal in relatedness.

Table 1

Results from Analyses of Variance of True and False Responses Treating Either Items (Items-analysis) or Subjects (Subjects-analysis) as a Random Variable while Averaging Across the Other

Trues

	Hi-PF	Med-PF	Lo-PF	F	df	p
Subjects-analysis	986	1045	1089	8.14	2,38	<.005
Items-analysis	979	1050	1112	2.41	2,36	n.s.

$\underline{\min F'}(2,59) = 1.86$, $\underline{p} > .10$

Falses

	High	Low	Anomalous	F	df	p
Subjects-analysis	1103	1248	998	38.58	2,38	<.001
Items-analysis	1136	1266	1001	10.24	2,36	<.001

$\underline{\min F'}(2,59) = 8.09$, $\underline{p} < .001$

Figure Captions

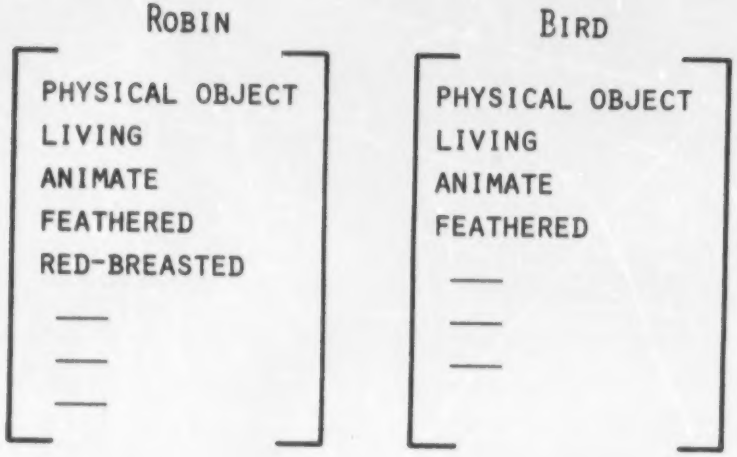
Figure 1. Two simple models of semantic memory--the Attribute and Hierarchical models.

Figure 2. An illustration of the Marker Search model. Lower-case letters designate the labels on relations.

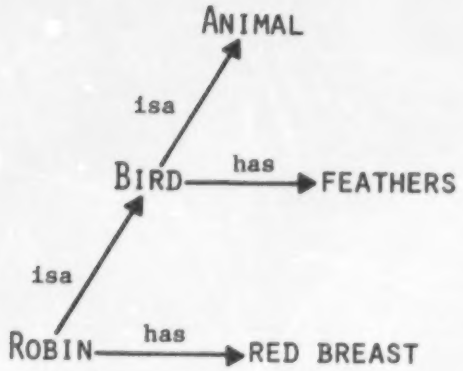
Figure 3. Illustrative cases of labeling contradictions.

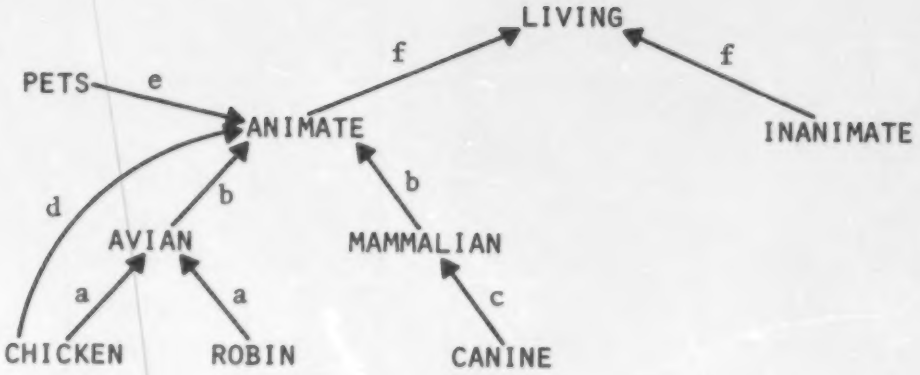
Figure 4. Illustration of labeling and short-cut paths.

ATTRIBUTE MODEL

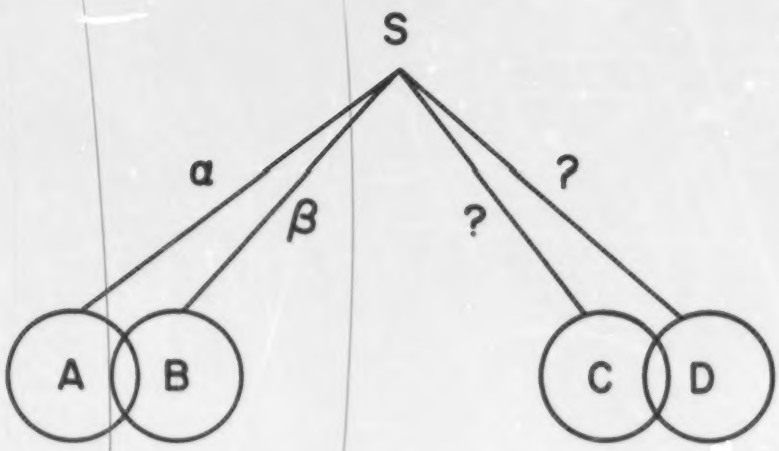


HIERARCHICAL NETWORK MODEL

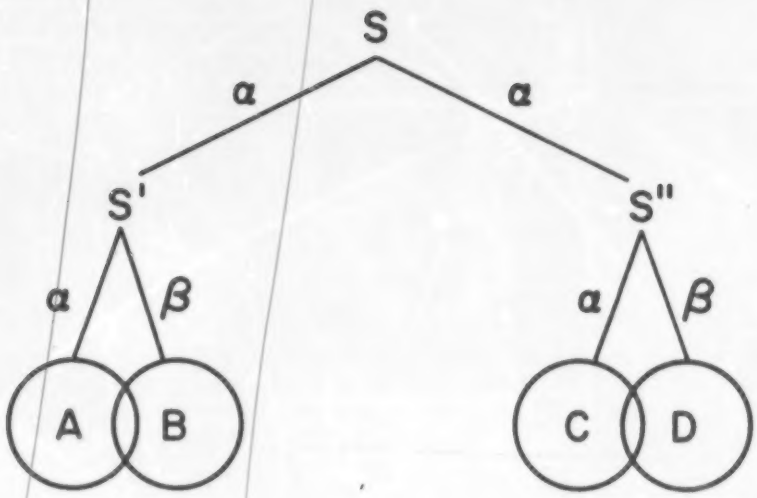



LIVING
f
f
PETS
e
ANIMATE
INANIMATE
d
b
b
AVIAN
MAMMALIAN
a
a
c
CHICKEN
ROBIN
CANINE

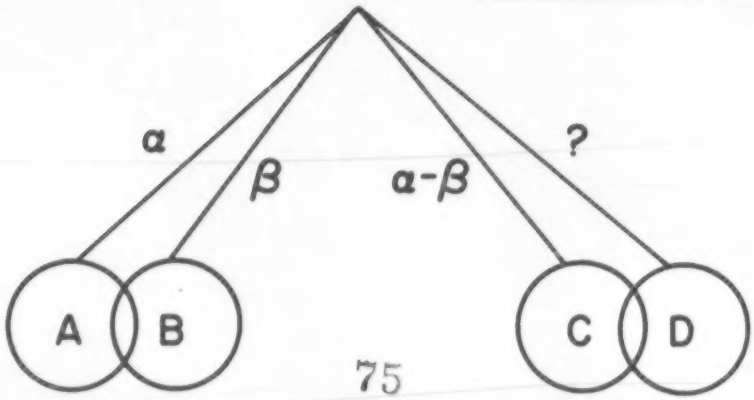
a

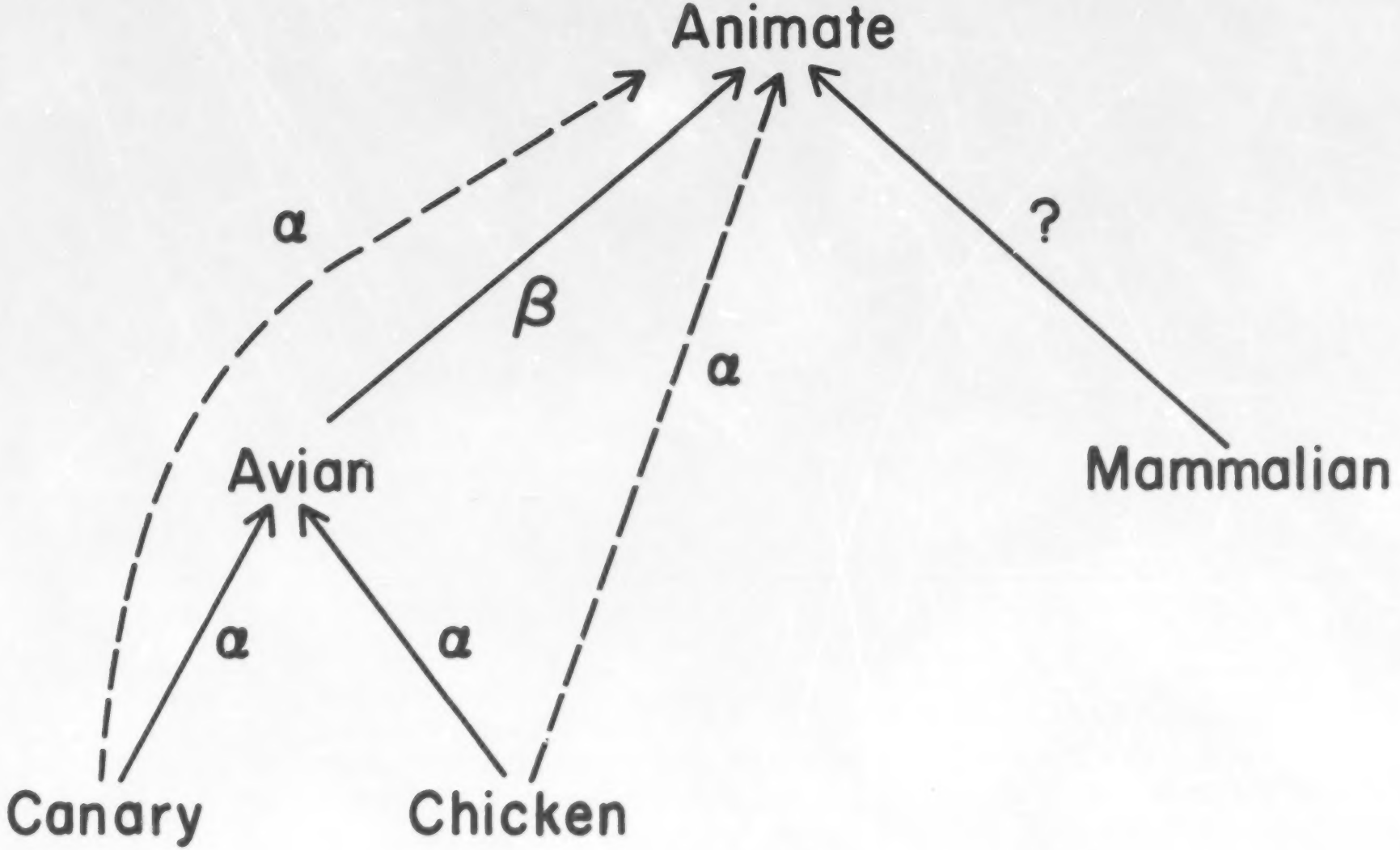


b



c




Animate
α
β
α
?
Avian
Mammalian
α
α
Canary
Chicken

CENTER FOR THE STUDY OF READING

READING EDUCATION REPORTS

No. 1: Durkin, D. Comprehension Instruction--Where Are You?, October 1977. (ERIC Document Reproduction Service No. ED 146 566, 14p., HC-$1.67, MF-$.83)

No. 2: Asher, S. R. Sex Differences in Reading Achievement, October 1977. (ERIC Document Reproduction Service No. ED 146 567, 30p., HC-$2.06, MF-$.83)

No. 3: Adams, M. J., Anderson, R. C., & Durkin, D. Beginning Reading: Theory and Practice, November 1977.

No. 4: Jenkins, J. R., & Pany, D. Teaching Reading Comprehension in the Middle Grades, January 1978.

No. 5: Bruce, B. What Makes a Good Story?, June 1978.

CENTER FOR THE STUDY OF READING

TECHNICAL REPORTS

* Available only through ERIC

*No. 1: Halff, H. M. Graphical Evaluation of Hierarchical Clustering Schemes, October 1975. (ERIC Document Reproduction Service No. ED 134 926, 11p., HC-$1.67, MF-$.83)

*No. 2: Spiro, R. J. Inferential Reconstruction in Memory for Connected Discourse, October 1975. (ERIC Document Reproduction Service No. ED 136 187, 81p., HC-$4.67, MF-$.83)

*No. 3: Goetz, E. T. Sentences in Lists and in Connected Discourse, November 1975. (ERIC Document Reproduction Service No. ED 134 927, 75p., HC-$3.50, MF-$.83)

*No. 4: Alessi, S. M., Anderson, T. H., & Biddle, W. B. Hardware and Software Considerations in Computer Based Course Management, November 1975. (ERIC Document Reproduction Service No. ED 134 928, 21p., HC-$1.67, MF-$.83)

*No. 5: Schallert, D. L. Improving Memory for Prose: The Relationship Between Depth of Processing and Context, November 1975. (ERIC Document Reproduction Service No. ED 134 929, 37p., HC-$2.06, MF-$.83)

*No. 6: Anderson, R. C., Goetz, E. T., Pichert, J. W., & Halff, H. M. Two Faces of the Conceptual Peg Hypothesis, January 1976. (ERIC Document Reproduction Service No. ED 134 930, 29p., HC-$2.06, MF-$.83)

*No. 7: Ortony, A. Names, Descriptions, and Pragmatics, February 1976. (ERIC Document Reproduction Service No. ED 134 931, 25p., HC-$1.67, MF-$.83)

*No. 8: Mason, J. M. Questioning the Notion of Independent Processing Stages in Reading, February 1976. (Journal of Educational Psychology, 1977, 69, 288-297)

*No. 9: Siegel, M. A. Teacher Behaviors and Curriculum Packages: Implications for Research and Teacher Education, April 1976. (ERIC Document Reproduction Service No. ED 134 932, 42p., HC-$2.06, MF-$.83)

*No. 10: Anderson, R. C., Pichert, J. W., Goetz, E. T., Schallert, D. L., Stevens, K. V., & Trollip, S. R. Instantiation of General Terms, March 1976. (ERIC Document Reproduction Service No. ED 134 933, 30p., HC-$2.06, MF-$.83)

*No. 11: Armbruster, B. B. Learning Principles from Prose: A Cognitive Approach Based on Schema Theory, July 1976. (ERIC Document Reproduction Service No. ED 134 934, 48p., HC-$.206, MF-$.83)

*No. 12: Anderson, R. C., Reynolds, R. E., Schallert, D. L., & Goetz, E. T. Frameworks for Comprehending Discourse, July 1976. (ERIC Document Reproduction Service No. ED 134 935, 33p., HC-$2.06, MF-$.83)

No. 13: Rubin, A. D., Bruce, B. C., & Brown, J. S. A Process-oriented language for Describing Aspects of Reading Comprehension, November 1976. (ERIC Document Reproduction Service No. ED 136 188, 41p., HC-$2.06, MF-$.83)

No. 14: Pichert, J. W., & Anderson, R. C. Taking Different Perspectives on a Story, November 1976. (ERIC Document Reproduction Service No. ED 134 936, 30p., HC-$2.06, MF-$.83)

No. 15: Schwartz, R. M. Strategic Processes in Beginning Reading, November 1976. (ERIC Document Reproduction Service No. ED 134 937, 19p., HC-$1.67, MF-$.83)

No. 16: Jenkins, J. R., & Pany, D. Curriculum Biases in Reading Achievement Tests, November 1976. (ERIC Document Reproduction Service No. ED 134 938, 24p., HC-$1.67, MF-$.83)

No. 17: Asher, S. R., Hymel, S., & Wigfield, A. Children's Comprehension of High- and Low-Interest Material and a Comparison of Two Cloze Scoring Methods, November 1976. (ERIC Document Reproduction Service No. ED 134 939, 32p., HC-$2.06, MF-$.83)

No. 18: Brown, A. L., Smiley, S. S., Day, J. D., Townsend, M. A. R., & Lawton, S. C. Intrusion of a Thematic Idea in Children's Comprehension and Retention of Stories, December 1976. (ERIC Document Reproduction Service No. ED 136 189, 39p., HC-$2.06, MF-$.83)

No. 19: Kleiman, G. M. The Prelinguistic Cognitive Basis of Children's Communicative Intentions, February 1977. (ERIC Document Reproduction Service No. ED 134 940, 51p., HC-$3.50, MF-$.83)

No. 20: Kleiman, G. M. The Effect of Previous Context on Reading Individual Words, February 1977. (ERIC Document Reproduction Service No. ED 134 941, 76p., HC-$4.67, MF-$.83)

No. 21: Kane, J. H., & Anderson, R. C. Depth of Processing and Interference Effects in the Learning and Remembering of Sentences, February 1977. (ERIC Document Reproduction Service No. ED 134 942, 29p., HC-$2.06, MF-$.83)

No. 22: Brown, A. L., & Campione, J. C. Memory Strategies in Learning: Training Children to Study Strategically, March 1977. (ERIC Document Reproduction Service No. ED 136 234, 54p., HC-$3.50, MF-$.83)

No. 23: Smiley, S. S., Oakley, D. D., Worthen, D., Campione, J. C., & Brown, A. L. Recall of Thematically Relevant Material by Adolescent Good and Poor Readers as a Function of Written Versus Oral Presentation, March 1977. (ERIC Document Reproduction Service No. ED 136 235, 23p., HC-$1.67, MF-$.83)

No. 24: Anderson, R. C., Spiro, R. J., & Anderson, M. C. Schemata as Scaffolding for the Representation of Information in Connected Discourse, March 1977. (ERIC Document Reproduction Service No. ED 136 236, 18p., HC-$1.67, MF-$.83)

No. 25: Pany, D., & Jenkins, J. R. Learning Word Meanings: A Comparison of Instructional Procedures and Effects on Measures of Reading Comprehension with Learning Disabled Students, March 1977. (ERIC Document Reproduction Service No. ED 136 237, 34p., HC-$2.06, MF-$.83)

No. 26: Armbruster, B. B., Stevens, R. J., & Rosenshine, B. Analyzing Content Coverage and Emphasis: A Study of Three Curricula and Two Tests, March 1977. (ERIC Document Reproduction Service No. ED 136 238, 22p., HC-$1.67, MF-$.83)

No. 27: Ortony, A., Reynolds, R. E., & Arter, J. A. Metaphor: Theoretical and Empirical Research, March 1977.

No. 28: Ortony, A. Remembering and Understanding Jabberwocky and Small-Talk, March 1977. (ERIC Document Reproduction Service No. ED 137 753, 36 p., HC-$2.06, MF-$.83)

No. 29: Schallert, D. L., Kleiman, G. M., & Rubin, A. D. Analysis of Differences Between Oral and Written Language, April 1977. (ERIC Document Reproduction Service No. ED 144 038, 33p., HC-$2.06, MF-$.83)

No. 30: Goetz, E. T., & Osborn, J. Procedures for Sampling Texts and Tasks in Kindergarten through Eighth Grade, April 1977. (ERIC Document Reproduction Service No. ED 146 565, 80p., HC-$4.67, MF-$.83)

No. 31: Nash-Webber, B. Anaphora: A Cross-Disciplinary Survey, April 1977. (ERIC Document Reproduction Service No. ED 144 039, 43p., HC-$2.06, MF-$.83)

No. 32: Adams, M. J., & Collins, A. A Schema-Theoretic View of Reading Comprehension, April 1977. (ERIC Document Reproduction Service No. ED 146 565, 80p., HC-$4.67, MF-$.83)

No. 33: Huggins, A. W. F. Syntactic Aspects of Reading Comprehension, April 1977. (ERIC Document Reproduction Service No. ED 142 972, 68p., HC-$3.50, MF-$.83)

No. 34: Bruce, B. C. Plans and Social Actions, April 1977.

No. 35: Rubin, A. D. Comprehension Processes in Oral and Written Language, April 1977.

No. 36: Nash-Webber, B., & Reiter, R. Anaphora and Logical Form: On Formal Meaning Representations for Natural Language, April 1977. (ERIC Document Reproduction Service No. ED 142 973, 42p., HC-$2.06, MF-$.83)

No. 37: Adams, M. J. Failures to Comprehend and Levels of Processing in Reading, April 1977.

No. 38: Woods, W. A. Multiple Theory Formation in High-Level Perception, April 1977. (ERIC Document Reproduction Service No. ED 144 020, 58p., HC-$3.50, MF-$.83)

No. 40: Collins, A., Brown, J. S., & Larkin, K. M. Inference in Text Understanding, December 1977.

No. 41: Anderson, R. C., & Pichert, J. W. Recall of Previously Unrecallable Information Following a Shift in Perspective, April 1977. (ERIC Document Reproduction Service No. ED 142 974, 37p., HC-$2.06, MF-$.83)

No. 42: Mason, J., Osborn, J., & Rosenshine, B. A Consideration of Skill Hierarchy Approaches to the Teaching of Reading, December 1977.

No. 43: Collins, A., Brown, A. L., Morgan, J. L., & Brewer, W. F. The Analysis of Reading Tasks and Texts, April 1977.

No. 44: McClure, E. Aspects of Code-Switching in the Discourse of Bilingual Mexican-American Children, April 1977. (ERIC Document Reproduction Service No. ED 142 975, 39p., HC-$2.06, MF-$.83)

No. 45: Schwartz, R. M. Relation of Context Utilization and Orthographic Automaticity in Word Identification, May 1977.

No. 46: Anderson, R. C., Stevens, K. C., Shifrin, Z., & Osborn, J. Instantiation of Word Meanings in Children, May 1977. (ERIC Document Reproduction Service No. ED 142 976, 22p., HC-$1.67, MF-$.83)

No. 47: Brown, A. L. Knowing When, Where, and How to Remember: A Problem of Metacognition, June 1977. (ERIC Document Reproduction Service No. ED 146 562, 152p., HC-$8.69, MF-$.83)

No. 48: Brown, A. L., & DeLoache, J. S. Skills, Plans, and Self-Regulation, July 1977. (ERIC Document Reproduction Service No. ED 144 040, 66p., HC-$3.50, MF-$.83)

No. 49: Goetz, E. T. Inferences in the Comprehension of and Memory for Text, July 1977.

No. 50: Anderson, R. C. Schema-Directed Processes in Language Comprehension, July 1977. (ERIC Document Reproduction Service No. ED 142 977, 33p., HC-$2.06, MF-$.83)

No. 51: Brown, A. L. Theories of Memory and the Problems of Development: Activity, Growth, and Knowledge, July 1977. (ERIC Document Reproduction Service No. ED 144 041, 59p., HC-$3.50, MF-$.83)

No. 52: Morgan, J. L. Two Types of Convention in Indirect Speech Acts, July 1977.

No. 53: Brown, A. L., Smiley, S. S., & Lawton, S. C. The Effects of Experience on the Selection of Suitable Retrieval Cues for Studying from Prose Passages, July 1977. (ERIC Document Reproduction Service No. ED 144 042, 30p., HC-$2.06, MF-$.83)

No. 54: Fleisher, L. S., & Jenkins, J. R. Effects of Contextualized and Decontextualized Practice Conditions on Word Recognition, July 1977. (ERIC Document Reproduction Service No. ED 144 043, 37p., HC-$2.06, MF-$.83)

No. 55: Jenkins, J. R., & Larson, K. Evaluating Error Correction Procedures for Oral Reading, June 1978.

No. 56: Anderson, T. H., Standiford, S. N., & Alessi, S. M. Computer Assisted Problem Solving in an Introductory Statistics Course, August 1977. (ERIC Document Reproduction Service No. ED 146 563, 26p., HC-$2.06, MF-$.83)

No. 57: Barnitz, J. Interrelationship of Orthography and Phonological Structure in Learning to Read, August 1977.

No. 58: Mason, J. M. The Role of Strategy in Reading in the Mentally Retarded, September 1977.

No. 59: Mason, J. M. Reading Readiness: A Definition and Skills Hierarchy from Preschoolers' Developing Conceptions of Print, September 1977.

No. 60: Spiro, R. J., & Esposito, J. J. Superficial Processing of Explicit Inferences in Text, December 1977.

No. 61: Spiro, R. J., & Smith, D. Distinguishing Sub-Types of Poor Comprehenders: Overreliance on Conceptual vs. Data-Driven Processes, April 1978.

No. 65: Brewer, W. F. Memory for the Pragmatic Implications of Sentences, October 1977. (ERIC Document Reproduction Service No. ED 146 564, 27p., HC-$2.06, MF-$.83)

No. 66: Brown, A. L., & Smiley, S. S. The Development of Strategies for Studying Prose Passages, October 1977.

No. 68: Stein, N. L., & Nezworski, T. The Effects of Organization and Instructional Set on Story Memory, January 1978.

No. 69: Stein, N. L. How Children Understand Stories: A Developmental Analysis, March 1978.

No. 76: Thieman, T. J., & Brown, A. L. The Effects of Semantic and Formal Similarity on Recognition Memory for Sentences in Children, November 1977.

No. 77: Nash-Webber, B. L. Inference in an Approach to Discourse Anaphora, January 1978.

No. 78: Gentner, D. On Relational Meaning: The Acquisition of Verb Meaning, December 1977.

No. 79: Royer, J. M. Theories of Learning Transfer, January 1978.

No. 80: Arter, J. A., & Jenkins, J. R. Differential Diagnosis-Prescriptive Teaching: A Critical Appraisal, January 1978.

No. 81: Shoben, E. J. Choosing a Model of Sentence Picture Comparisons: A Reply to Catlin and Jones, February 1978.

No. 82: Steffensen, M. S. Bereiter and Engelmann Reconsidered: The Evidence from Children Acquiring Black English Vernacular, March 1978.

No. 83: Reynolds, R. E., Standiford, S. N., & Anderson, R. C. Distribution of Reading Time when Questions are Asked about a Restricted Category of Text Information, April 1978.

No. 84: Baker, L. Processing Temporal Relationships in Simple Stories: Effects of Input Sequence, April 1978.

No. 85: Mason, J. M., Knisely, E., & Kendall, J. Effects of Polysemous Words on Sentence Comprehension, May 1978.

No. 86: Anderson, T. H., Wardrop, J. L., Hively, W., Muller, K. E., Anderson, R. I., Hastings, C. N., & Frederiksen, J. Development and Trial of a Model for Developing Domain Referenced Tests of Reading Comprehension, May 1978.

No. 87: André, M. E. D. A., & Anderson, T. H. The Development and Evaluation of a Self-Questioning Study Technique, June 1978.

No. 88: Bruce, B., & Newman, D. Interacting Plans, June 1978.

No. 89: Bruce, B., Collins, A., Rubin, A. D., & Gentner, D. A Cognitive Science Approach to Writing, June 1978.

No. 90: Asher, S. T. Referential Communication, June 1978.

No. 91: Royer, J. M., & Cunningham, D. J. On the Theory and Measurement of Reading Comprehension, June 1978.

No. 92: Mason, J. M., & Kendall, J. R. Facilitating Reading Comprehension through Text Structure Manipulation, June 1978.

No. 93: Ortony, A., Schallert, D. L., Reynolds, R. E., & Antos, S. J. Interpreting Metaphors and Idioms: Some Effects of Context on Comprehension, July 1978.

No. 94: Brown, A. L., Campione, J. C., & Barclay, C. R. Training Self-checking Routines for Estimating Test Readiness: Generalization from List Learning to Prose Recall, July 1978.

No. 95: Reichman, R. Conversational Coherency, July 1978.

No. 96: Wigfield, A., & Asher, S. R. Age Differences in Children's Referential Communication Performance: An Investigation of Task Effects, July 1978.

No. 97: Steffensen, M. S., Jogdeo, C., & Anderson, R. C. A Cross-Cultural Perspective on Reading Comprehension, July 1978.

No. 98: Green, G. M. Discourse Functions of Inversion Constructions, July 1978.

No. 99: Asher, S. R. Influence of Topic Interest on Black Children and White Children's Reading Comprehension, July 1978.

No. 100: Jenkins, J. R., Pany, D., & Schreck, J. Vocabulary and Reading Comprehension: Instructional Effects, August 1978.

No. 101: Shoben, E. J., Rips, L. J., & Smith, E. E. Issues in Semantic Memory: A Response to Glass and Holyoak, August 1978.